基于深度学习的低照度图像与视频增强综述

Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, *Senior Member, IEEE*, Jinwei Gu, *Senior Member, IEEE*, and Chen Change Loy, *Senior Member, IEEE*

摘要— 低照度图像增强（LLIE）的目的是改善在光照不足的环境中拍摄的图像的感知或可解释性。这一领域的最新进展主要是基于深度学习的解决方案，其中采用了各种学习策略、网络结构、损失函数、训练数据等。在本文中，我们提供了一个全面的调查，涵盖了从算法分类到未解决的公开问题等各个方面。为了考察现有方法的通用性，我们提出了一个大规模的低照度图像和视频数据集，其中的图像和视频是由各种手机摄像头在不同的光照条件下拍摄的。此外，我们首次提供了一个统一的，涵盖了许多流行 LLIE 方法的在线平台，其结果可以通过一个用户友好的网页展示出来。除了在公开的和我们提出的数据集上对现有的方法进行定性和定量评估外，我们还验证了它们在黑暗中的人脸检测性能。这项调查连同提议的数据集和在线平台可以作为未来研究的参考来源，并促进这一研究领域的发展。上述的平台中的收集的算法、数据集和评估指标皆是公开的，并将定期更新。项目网址：https://www.mmlab-ntu.com/project/lliv_survey/index.html.

Index Terms—image and video restoration, low-light image dataset, low-light image enhancement platform, computational photography.

1 简介

由于不可避免的环境和/或技术限制，如照明不足和曝光时间有限，图像往往是在次优的照明条件下拍摄的，受到背光、非均匀光照和弱光的干扰。这类图像的美学质量会受到影响，而且对于高层的任务，如物体跟踪、识别和检测，信息的传输也不尽人意。图 1显示了由次优照明条件引起的退化的一些例子。

低照度增强在不同领域享有广泛的应用，包括视觉监控、自动驾驶和计算摄影。特别是，智能手机摄影已经变得普遍且流行。受限于手机相机光圈的大小、实时处理的需求以及内存的限制，在昏暗的环境中用智能手机的相机拍照尤其具有挑战性。在这种应用中，增强低光图像和视频是一个值得探索的研究领域。

传统的低光增强方法包括基于直方图均衡的方法 [1], [2] 和基于 Retinex 模型的方法 [3], [4], [5], [6], [7], [8], [9], [10]。后者受到的关注相对较多。一个典型的基于 Retinex 模型的方法通过某种先验或正则化将低照度图像分解为反射分量和照明分量。而被估测的反射分量被视为增强的结果。

这种方法有一些局限性：**1)** 将反射分量视为增强结果的理想假设并不总是成立，特别是考虑到各种照明特性，这可能导致不现实的增强，如细节的损失和色彩的扭曲，**2)** Retinex 模型中通常忽略了噪声，因此噪声在增强结果中被保留或被放大，**3)** 找到一个有效的先验或正则化是具有挑战性的。不准确的先验或正则化可能导致增强后的结果出现伪影和颜色偏差，以及 **4)** 由于其复杂的优化过程，运行时间相对较长。

(a) 背光 (b) 非均匀照明 (c) 弱光 (d) 极弱光 (e) 多色光 (f) 噪点

图 1. 在次优照明条件下拍摄的图像实例。这些图像存在着场景内容被埋没、对比度降低、强噪点和色彩不准确的问题。

近年来，自第一项开创性的工作 [11] 以来，基于深度学习的 LLIE 取得了令人瞩目的成功。与传统方法相比，基于深度学习的解决方案具有更好的准确性、鲁棒性和速度，因此近年来吸引了越来越多的关注。图 2显示了基于深度学习的 LLIE 方法的一个简明的里程碑。如图所示，自 2017 年以来，基于深度学习的解决方案的数量逐年增加。这些解决方案中使用的学习策略包括监督学习（SL）、强化学习（RL）、无监督学习（UL）、零次学习（ZSL）和半监督学习（SSL）。请注意，我们在图 2中只报告了一些代表性的方法。事实上，从 2017 年到 2020 年，基于深度学习的方法有 100 多篇论文，比传统方法的总数还多。此外，尽管一些一般的照片增强方法 [45], [46], [47], [48], [49], [50], [51], [52], [53] 可以在一定程度上提高图像的亮度，但我们在本调查中省略了这些方法，因为它们不是为了处理多样化的低光条件。我们专注于基于深度

C. Li and C. C. Loy are with the S-Lab, Nanyang Technological University (NTU), Singapore (e-mail: chongyi.li@ntu.edu.sg and ccloy@ntu.edu.sg).
C. Guo, L. Han, and M. M. Cheng are with the College of Computer Science, Nankai University, Tianjin, China (e-mail: guochunle@nankai.edu.cn, lhhan@mail.nankai.edu.cn, and cmm@nankai.edu.cn).
J. Jiang and J. Gu are with the SenseTime (e-mail: jiangjun@sensebrain.site and gujinwei@sensebrain.site).
C. Li and C. Guo contribute equally.
C. C. Loy is the corresponding author.

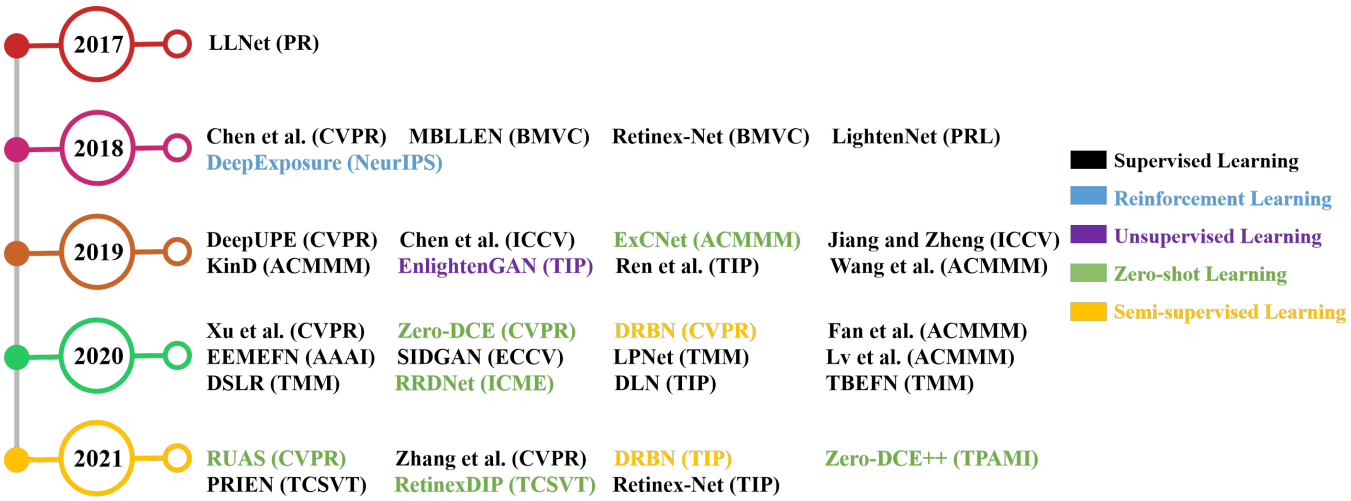


图 2. 基于深度学习的低照度图像和视频增强方法的简明里程碑。基于监督学习的方法: LLNet [11], Chen et al. [12], MBLLEN [13], Retinex-Net [14], LightenNet [15], SCIE [16], DeepUPE [17], Chen et al. [18], Jiang and Zheng [19], Wang et al. [20], KinD [21], Ren et al. [22], Xu et al. [23], Fan et al. [24], Lv et al. [25], EEMEFN [26], SIDGAN. [27], LPNet [28], DLN [29], TBEFN [30], DSLR [31], Zhang et al. [32], PRIEN [33], and Retinex-Net [34]. **基于强化学习的方法**: DeepExposure [35]. **基于无监督学习的方法**: EnlightenGAN [36]. **基于零次学习的方法**: ExCNet [37], Zero-DCE [38], RRDNet [39], Zero-DCE++ [40], RetinexDIP [41], and RUAS [42]. **基于半监督学习的方法**: DRBN [43] and DRBN [44].

学习的且专门为低照度图像和视频增强而开发的解决方案。

尽管深度学习在 LLIE 的研究中占主导地位，但对基于深度学习的解决方案还缺乏深入全面的调查。目前有两篇关于 LLIE 的回顾 [54], [55]。Wang 等人 [54] 主要回顾了传统 LLIE 的方法，而我们系统地、全面地回顾基于深度学习的 LLIE 的最新进展。Liu 等人 [55] 回顾了现有的 LLIE 算法，衡量了不同方法的机器视觉性能，提供了一个用于底层和高层视觉增强任务的低照度图像数据集，并开发了一个增强的人脸检测器。相比之下，我们的调查从不同方面回顾了低照度图像和视频增强，且具有以下独特的特点：**1)** 我们的工作主要集中在基于深度学习的低照度图像和视频增强的最新进展上，我们从各个方面进行了深入的分析和讨论，包括学习策略、网络结构、损失函数、训练数据集、测试数据集、评价指标、模型大小、推理速度、增强性能等。因此，本次调查的中心是深度学习及其在低照度图像和视频增强中的应用。**2)** 我们提出了一个数据集，其中包含由不同手机摄像头在不同光照条件下拍摄的图像和视频，以评估现有方法的通用性。这个新的、具有挑战性的数据集是对现有低照度图像和视频增强数据集的补充，因为在这个研究领域缺乏这样的数据集。此外，据我们所知，我们的这项工作是第一个在这种数据上对比基于深度学习的各种低照度图像增强方法性能的调查。**3)** 我们提供了一个在线平台，涵盖了许多流行的基于深度学习的低照度图像增强的方法，其结果可以通过一个用户友好的网页产生。通过我们的平台，即使没有 GPU 的用户也可以在线评测任何输入图像在使用不同增强方法后的结果。这加快了这个研究领域的发展，并有助于启发新的研究。我们希望我们的调查可以提供新的见解和灵感，以促进对基于深度学习的 LLIE 的理解，促进对提出的开放问题的研究，并加快这一研究领域的发展。

2 基于深度学习的 LLIE

2.1 问题定义

我们首先给出一个基于深度学习的 LLIE 问题的通用表述。对于宽度为 W、高度为 H 的低照度图像 $I \in \mathbb{R}^{W\times H\times 3}$，该过程可以被建模为：

$$\widehat{R} = \mathcal{F}(I;\theta), \tag{1}$$

其中 $\widehat{R} \in \mathbb{R}^{W\times H\times 3}$ 是增强的结果，$\mathcal{F}$ 代表具有可训练参数 θ 的网络。深度学习的目的是找到最佳网络参数 $\widehat{\theta}$ 使误差最小，即：

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}(\widehat{R}, R), \tag{2}$$

其中 $R \in \mathbb{R}^{W\times H\times 3}$ 是 ground truth，损失函数 $\mathcal{L}(\widehat{R}, R)$ 驱动网络的优化。在网络训练的过程中，可以使用各种损失函数，如监督损失和无监督损失。更多细节将在第3节介绍。

2.2 学习策略

根据不同的学习策略，我们将现有的 LLIE 方法分为监督学习、强化学习、无监督学习、零次学习和半监督学习。图3展示了从不同角度进行的统计学分析。在接下来的内容中，我们将回顾每种策略的一些代表性方法。

监督学习. 对于基于监督学习的 LLIE 方法，又分为端到端、基于深度 Retinex 和现实数据驱动的方法。

第一个基于深度学习的 LLIE 方法 LLNet [11] 采用了叠加稀疏去噪自动编码器 [56] 的变体，同时对低光图像进行增亮和去噪。这项开创性的工作激发了 LLIE 中端到端网络的使用。Lv 等人 [13] 提出了一个端到端多分支增强网络（MBLLEN）。MBLLEN 通过一个特征提取模块、一个增强模块和一个融合模块来提取有效的特征表示，提高了 LLIE 的性能。这位学者 [25] 还提出了其他三个子网络，包括 Illumination-Net，Fusion-Net，和 Restoration-Net 以进一

步提高性能。Ren 等人 [22] 设计了一个更复杂的端到端网络，包括一个用于图像内容增强的编码器-解码器网络和一个用于图像边缘增强的递归神经网络。与 Ren 等人 [22] 类似，Zhu 等人 [26] 提出了一种叫做 EEMEFN 的方法。EEMEFN 包括两个阶段：多曝光融合和边缘增强。多重曝光融合网络，TBEFN [30]，是为 LLIE 提出的。TBEFN 在两个分支中估算一个传递函数，其中可以得到两个增强结果。最后，采用一个简单的平均方案来融合这两个图像，并通过一个细化单元进一步细化结果。此外，pyramid network (LPNet) [28]，residual network [29]，和 Laplacian pyramid [31] (DSLR) 被引入 LLIE 中。这些方法可以在 LLIE 中通过常用的端到端网络结构，学习有效和高效地整合特征表示。基于对噪声在不同频率层表现出不同的对比度的观察，Xu 等人 [57] 提出了一个基于频率的分解和增强网络。该网络在低频层抑制噪声的情况下恢复图像内容，同时推断出高频层的细节。最近，有人提出了一个渐进式递归低照度图像增强网络 [33] 它使用一个递归单元来逐步增强输入图像。为了解决处理弱光视频时的时间不稳定性，Zhang 等人 [32] 提出从单一图像中学习和推理的动态场，然后强化时间一致性。

与在端到端网络中直接学习增强结果相比，由于物理上可解释的 Retinex 理论 [58], [59]，基于深度 Retinex 的方法在大多数情况下享有更好的增强性能。基于深度 Retinex 的方法通常通过专门的子网络分别增强光照度成分和反射度成分。Retinex-Net [14] 被提出，它包括一个将输入图像分割成与光无关的反射率和结构感知的平滑照明的 Decom-Net 模块，和一个调整照度图以实现低光增强的 Enhance-Net 模块。最近，Retinex-Net [14] 通过增加新的约束条件和先进的网络设计进行了扩展，以获得更好的增强性能 [34]。为了减少计算负担，Li 等人 [15] 提出了一种用于弱光图像增强的轻量级模型 LightenNet，它只由四层组成。LightenNet 将弱光图像作为输入，然后估算其照度图。基于 Retinex 理论 [58], [59], 增强的图像是通过将输入的图像除以照度图得到的。为了准确估计照度图，Wang 等人 [60] 通过他们提出的 DeepUPE 网络提取全局和局部特征来学习图像到照度的映射关系。Zhang 等人 [21] 分别开发了三个子网络，用于图层分解、反射率恢复和光照度调整，称为 KinD。此外，作者还通过一个多尺度光照注意力模块缓解了 KinD [21] 的结果中留下的视觉缺陷。改进后的 KinD 被称为 KinD++ [61]。为了解决基于深层 Retinex 方法中遗漏噪声的问题，Wang 等人 [20] 提出了一个渐进式 Retinex 网络，其中一个 IM 网络估算光照度，一个 NM 网络估算噪声水平。这两个子网络以一种渐进的机制工作，直到获得稳定的结果。Fan 等人 [24] 将语义分割和 Retinex 模型结合起来，以进一步提高实际情况下的增强性能。其核心思想是利用图像语义的先验信息来引导光照成分和反射成分的增强。

尽管有些方法可以达到不错的性能，但由于使用了合成训练数据，它们在真实的低照度情况下显示出较差的泛化能力。为了解决这个问题，一些工作试图合成更真实的训练数据或获取真实世界中的数据。Cai 等人 [16] 建立了一个多曝光图像数据集，其中不同曝光度的低对比度图像有其相应的高质量参考图像。每张高质量的参考图像都是通过主观选择不同方法增强的 13 个结果中的最佳输出而获得的。此外，在建立的数据集上，他们训练了一个频率分解网络，并通过两阶段结构分别增强高频层和低频层。Chen 等人 [12] 收集了一个真实的低光图像数据集（SID），并训练 U-Net [62] 来学习从低光原始数据到相应的长曝光高质量参考图像的映射。此外，Chen 等人 [18] 将 SID 数据集扩展到低光视频（DRV）。DRV 包含静态视频和相应的长曝光的 ground truths。为了保证处理动态场景视频的泛化能力，他们提出了一个孪生神经网络。为了增强黑暗中的移动物体，Jiang 和 Zheng [19] 设计了一个共轴光学系统来捕捉时间上同步和空间上一致的低光和高光视频组（SMOID）。与 DRV 视频数据集 [18] 不同，SMOID 视频数据集包含动态场景。为了学习从原始低光视频到光线充足视频的映射，他们提出了一个基于 U-Net 的三维网络。考虑到以前的低光视频数据集的局限性，如 DRV 数据集 [18] 只包含统计视频和 SMOID 数据集 [19] 的 179 个视频组，Triantafyllidou 等人 [27] 提出一个低照度视频合成管线，称为 SIDGAN。SIDGAN 可以通过半监督的双 CycleGAN 与中间域映射产生动态视频数据（raw 到 RGB）。从 Vimeo-90K 数据集 [63] 中收集来的真实世界的视频被用来训练这个管线。低照度原始视频数据和相应的长曝光图像是从 DRV 数据集 [18] 中提取的。通过合成的训练数据，这项工作采用了与 Chen 等人 [12] 相同的 U-Net 网络进行低光视频增强。

强化学习. 在没有配对训练数据的情况下，Yu 等人 [35] 用强化对抗学习来学习照片的曝光，其命名为 DeepExposure。具体来说，输入的图像首先根据曝光度被分割成子图像。对于每个子图像，局部曝光是由策略网络基于强化学习依次学习的。奖励评价函数由对抗性学习来逼近。最后，每个局部曝光被用于修饰输入，从而获得不同曝光下的多个修饰图像。最终的结果是通过融合这些图像来实现的。

无监督学习. 在成对的数据上训练一个深度模型可能会导致过度拟合从而限制泛化能力。为了解决这个问题，一种名为 EnligthenGAN [36] 的无监督学习方法被提出。EnlightenGAN 采用注意力引导的 U-Net [62] 作为生成器，并使用全局-局部判别器来确保增强的结果看起来像真实的正常光线图像。除了全局和局部对抗性损失外，还提出了全局和局部自我特征保留损失，以保留增强前后的图像内容。这是稳定训练这种单路径生成对抗网络（GAN）结构的一个关键点。

零次学习. 监督学习、强化学习和无监督学习方法要么泛化能力有限，要么受到不稳定训练的影响。为了弥补这些问题，我们提出了零次学习法，仅从测试图像中学习增强。请注意，在底层视觉任务中使用零次学习的概念是为了强调该方法不需要配对或非配对的训练数据，这与它在高层视觉任务中的定义不同。Zhang 等人 [37] 提出了一种零次学习方法，称为

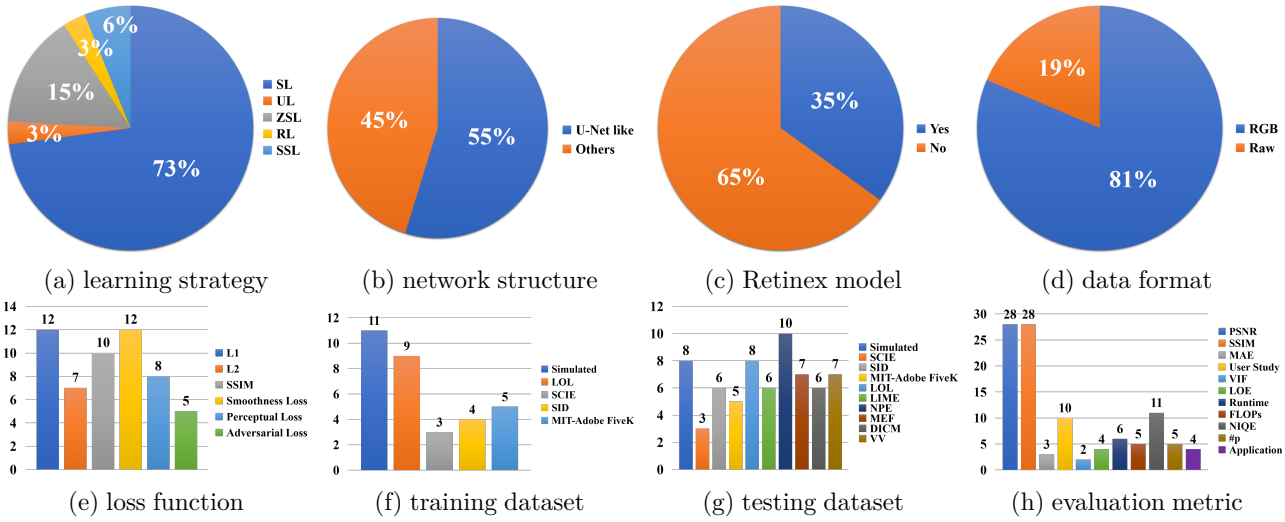


图 3. 基于深度学习的 LLIE 方法的统计分析，包括学习策略、网络特征、Retinex 模型、数据格式、损失函数、训练数据集、测试数据集和评价指标。可放大图片看详情。

ExCNet，用于背光图像修复。其首先使用一个网络来估算最适合输入图像的 S 曲线。一旦 S 曲线被估算出来，输入图像就通过引导滤波器 [64] 被分离成一个基础层和一个细节层。然后通过估算的 S 曲线来调整基础层。最后，Weber 对比度 [65] 被用来融合细节层和调整后的基础层。为了训练 ExCNet，作者将损失函数表述为一个基于块的能量最小化问题。Zhu 等人 [39] 提出一个三分支的 CNN，称为 RRDNet，用于曝光不足的图像修复。RRDNet 通过迭代最小化专门设计的损失函数，将输入图像分解为光照度、反射率和噪声。为了驱动零次学习，其提出了 Retinex 重建损失、纹理增强损失和照度引导的噪声估算损失的损失函数组合。Zhao 等人 [41] 通过神经网络进行 Retinex 分解，然后使用基于 Retinex 的 RetinexDIP 模型来增强低照度图像。受 Deep Image Prior（DIP）[66] 的启发，RetinexDIP 通过随机采样的白噪声生成输入图像的反射分量和照度分量，其中照度平滑度等分量特征相关损失被用于训练。Liu 等人 [42] 提出了一种受 Retinex 启发的 LLIE 的解卷方法，其中合作架构搜索被用来发现基本块的轻量级先验架构，非参考损失被用来训练网络。与基于图像重建的方法 [11], [13], [14], [21], [22], [31], [61] 不同，一种深度曲线估算网络 Zero-DCE [38] 被提出。Zero-DCE 将光增强视为一项针对图像的曲线估算任务，它将低光图像作为输入，并产生高阶曲线作为其输出。这些曲线被用来对输入的动态范围进行像素级的调整，以获得增强的图像。此外，还提出了一个快速和轻量级的版本，称为 Zero-DCE++ [40]。这种基于曲线的方法在训练期间不需要任何配对或非配对的数据。它们通过一组非参考损失函数实现零参考学习。此外，与需要高计算资源的基于图像重建的方法不同，图像到曲线的映射只需要轻量级的网络，因此实现了快速推算。

半监督学习. 为了结合监督学习和无监督学习的优势，近年来有人提出了半监督学习。Yang 等人 [43] 提出了一个半监督的深度递归波段网络（DRBN）。DRBN 首先在监督学习下恢复增强图像的线性波段表示，然后通过基于无监督对抗学习的可学习线性变换对给定波段进行重新组合，获得改进的波段表示。通过引入长短时记忆（LSTM）网络和在美学视觉分析数据集上预训练的图像质量评估网络，DRBN 得到了扩展，实现了更好的增强性能 [44].

通过观察图 3(a)，我们可以发现，在基于深度学习的 LLIE 方法中，监督学习是主流，其中比例达到 73%。这是因为当使用成对的训练数据，如 LOL [14], SID [12] 和各种低光/常光图像合成方法时，监督学习相对容易。然而，基于监督学习的方法面临一些挑战: **1)** 收集大规模的配对数据集，涵盖不同的现实世界的低光条件是困难的，**2)** 合成的低光图像不能准确地代表现实世界的照度条件，如空间变化的照明和不同程度的噪声，**3)** 在配对数据上训练一个深度模型可能导致对现实世界不同照度的图像的泛化能力弱。

因此，一些方法采用了无监督学习、强化学习、半监督学习和零次学习来绕过监督学习面临的局限。虽然这些方法取得了极具竞争性的性能，但它们仍然受到一些限制：**1)** 对于无监督学习/半监督学习方法，如何实现稳定的训练，避免颜色偏差，并建立跨域信息的关系，是对当前方法的挑战；**2)** 对于强化学习方法，设计有效的奖励机制和实现高效稳定的训练是错综复杂的；**3)** 对于零点学习方法，当要考虑到颜色保存、伪影去除和梯度反传播时，非参考损失的设计是不容易的。

3 技术回顾与讨论

在本节中，我们首先总结了表1中具有代表性的基于深度学习的 LLIE 方法，然后分析和讨论它们的技术特点。

3.1 网络结构

现有模型中使用了多种网络结构和设计，从基本的 U-Net、pyramid 网络、多阶段网络到频率分解网络。在分析了图3(b)

表 1
基于深度学习的代表性方法的基本特征总结。“Retinex”表示模型是否是基于 Retinex。“simulated”表示测试数据的模拟方法与合成训练数据的模拟方法相同。“self-selected”表示作者选择的真实世界的图像。“#P”表示训练参数量。“-”表示该项目不存在或未在原文中注明。

	Method	Learning	Network Structure	Loss Function	Training Data	Testing Data	Evaluation Metric	Format	Platform	Retinex
2017	LLNet [11]	SL	SSDA	SRR loss	simulated by Gamma Correction & Gaussian Noise	simulated self-selected	PSNR SSIM	RGB	Theano	
2018	LightenNet [15]	SL	four layers	L_2 loss	simulated by random illumination values	simulated self-selected	PSNR MAE SSIM User Study	RGB	Caffe MATLAB	✓
	Retinex-Net [14]	SL	multi-scale network	L_1 loss smoothness loss invariable reflectance loss	LOL simulated by adjusting histogram	self-selected	-	RGB	TensorFlow	✓
	MBLLEN [13]	SL	multi-branch fusion	SSIM loss region loss perceptual loss	simulated by Gamma Correction & Poisson Noise	simulated self-selected	PSNR SSIM AB VIF LOE TOMI	RGB	TensorFlow	
	SCIE [16]	SL	frequency decomposition	L_2 loss L_1 loss SSIM loss	SCIE	SCIE	PSNR FSIM Runtime FLOPs	RGB	Caffe MATLAB	
	Chen et al. [12]	SL	U-Net	L_1 loss	SID	SID	PSNR SSIM	raw	TensorFlow	
	Deepexposure [35]	RL	policy network GAN	deterministic policy gradient adversarial loss	MIT-Adobe FiveK	MIT-Adobe FiveK	PSNR SSIM	raw	TensorFlow	
2019	Chen et al. [18]	SL	siamese network	L_1 loss self-consistency loss	DRV	DRV	PSNR SSIM MAE	raw	TensorFlow	
	Jiang and Zheng [19]	SL	3D U-Net	L_1 loss	SMOID	SMOID	PSNR SSIM MSE	raw	TensorFlow	
	DeepUPE [60]	SL	illumination map	L_1 loss color loss smoothness loss	retouched image pairs	MIT-Adobe FiveK	PSNR SSIM User Study	RGB	TensorFlow	✓
	KinD [21]	SL	three subnetworks U-Net	reflectance similarity loss illumination smoothness loss mutual consistency loss L_1 loss L_2 loss SSIM loss texture similarity loss illumination adjustment loss	LOL	LOL LIME NPE MEF	PSNR SSIM LOE NIQE	RGB	TensorFlow	✓
	Wang et al. [20]	SL	two subnetworks pointwise Conv	L_1 loss	simulated by camera imaging model	IP100 FNF38 MPI LOL NPE	PSNR SSIM NIQE	RGB	Caffe	✓
	Ren et al. [22]	SL	U-Net like network RNN dilated Conv	L_2 loss perceptual loss adversarial loss	MIT-Adobe FiveK with Gamma correction & Gaussion noise	simulated self-selected DPED	PSNR SSIM Runtime	RGB	Caffe	
	EnlightenGAN [36]	UL	U-Net like network	adversarial loss self feature preserving loss	unpaired real images	NPE LIME MEF DICM VV BBD-100K ExDARK	User Study NIQE Classification	RGB	PyTorch	
	ExCNet. [37]	ZSL	fully connected layers	energy minimization loss	real images	$IE_{ps}D$	User Study CDIQA LOD	RGB	PyTorch	
2020	Zero-DCE [38]	ZSL	U-Net like network	spatial consistency loss exposure control loss color constancy loss illumination smoothness loss	SICE	SICE NPE LIME MEF DICM VV DARK FACE	User Study PI PNSR SSIM MAE Runtime Face detection	RGB	PyTorch	
	DRBN [43]	SSL	recursive network	SSIM loss perceptual loss adversarial loss	LOL images selected by MOS	LOL	PSNR SSIM SSIM-GC	RGB	PyTorch	
	Lv et al. [25]	SL	U-Net like network	Huber loss SSIM loss perceptual loss illumination smoothness loss	simulated by a retouching module	LOL SICE DeepUPE	User Study PSNR SSIM VIF LOE NIQE #P Runtime Face detection	RGB	TensorFlow	✓
	Fan et al. [24]	SL	four subnetworks U-Net like network feature modulation	mutual smoothness loss reconstruction loss illumination smoothness loss cross entropy loss consistency loss SSIM loss gradient loss ratio learning loss	simulated by illumination adjustment, slight color distortion, and noise simulation	simulated self-selected	PSNR SSIM NIQE	RGB	-	✓
	Xu et al. [57]	SL	frequency decomposition U-Net like network	L_2 loss perceptual loss	SID in RGB	SID in RGB self-selected	PSNR SSIM	RGB	PyTorch	
	EEMEFN [26]	SL	U-Net like network edge detection network	L_1 loss weighted cross-entropy loss	SID	SID	PSNR SSIM	raw	TensorFlow PaddlePaddle	
	DLN [29]	SL	residual learning interactive factor back projection network	SSIM loss total variation loss	simulated by illumination adjustment, slight color distortion, and noise simulation	simulated LOL	User Study PSNR SSIM NIQE	RGB	PyTorch	
	LPNet [28]	SL	pyramid network	L_1 loss perceptual loss luminance loss	LOL SID in RGB MIT-Adobe FiveK	LOL SID in RGB MIT-Adobe FiveK MEF NPE DICM VV	PSNR SSIM NIQE #P FLOPs Runtime	RGB	PyTorch	
	SIDGAN [27]	SL	U-Net	CycleGAN loss	SIDGAN	SIDGAN	PSNR SSIM TPSNR TSSIM ATWE	raw	TensorFlow	
	RRDNet [39]	ZSL	three subnetworks	retinex reconstruction loss texture enhancement loss noise estimation loss	-	NPE LIME MEF DICM	NIQE CPCQI	RGB	PyTorch	✓
	TBEFN [30]	SL	three stages U-Net like network	SSIM loss perceptual loss smoothness loss	SCIE LOL	SCIE LOL DICM MEF NPE VV	PSNR SSIM NIQE Runtime #P FLOPs	RGB	TensorFlow	✓
	DSLR [31]	SL	Laplacian pyramid U-Net like network	L_2 loss Laplacian loss color loss	MIT-Adobe FiveK	MIT-Adobe FiveK self-selected	PSNR SSIM NIQMC NIQE BTMQI CaHDC	RGB	PyTorch	
2021	RUAS [42]	ZSL	neural architecture search	cooperative loss similar loss total variation loss	LOL MIT-Adobe FiveK	LOL MIT-Adobe FiveK	PSNR SSIM Runtime #P FLOPs	RGB	PyTorch	✓
	Zhang et al. [32]	SL	U-Net	L_1 loss consistency loss	simulated by illumination adjustmentand noise simulation	simulated self-selected	User Study PSNR SSIM AB MABD WE	RGB	PyTorch	
	Zero-DCE++ [40]	ZSL	U-Net like network	spatial consistency loss exposure control loss color constancy loss illumination smoothness loss	SICE	SICE NPE LIME MEF DICM VV DARK FACE	User Study PI PNSR SSIM #P MAE Runtime Face detection FLOPs	RGB	PyTorch	
	DRBN [44]	SSL	recursive network	perceptual loss detail loss quality loss	LOL	LOL	PSNR SSIM SSIM-GC	RGB	PyTorch	
	Retinex-Net [34]	SL	three subnetworks	L_1 loss L_2 loss SSIM loss total variation loss	LOL simulated by adjusting histogram	LOL simulated NPE DICM VV	PNSR SSIM UQI OSS User Study	RGB	PyTorch	✓
	RetinexDIP [41]	ZSL	encoder-decoder networks	reconstruction loss illumination-consistency loss reflectnce loss illumination smoothness loss	-	DICM, ExDark Fusion LIME NASA NPE VV	NIQE NIQMC CPCQI	RGB	PyTorch	✓
	PRIEN [33]	SL	recursive network	SSIM loss	MEF LOL simulated by adjusting histogram	LOL LIME NPE MEF VV	PNSR SSIM LOE TMQI	RGB	PyTorch	

之后，可以看出 U-Net 和类 U-Net 网络是 LLIE 中主要采用的网络结构。这是因为 U-Net 可以有效地整合多尺度特征，并同时利用低级和高级特征。这些特点对于实现令人满意的低照度增强是至关重要的。然而，目前的 LLIE 网络结构可能忽略了一些关键问题：**1)** 在经过几个卷积层后，由于像素值较小，极低照度图像的梯度在梯度反向传播中可能消失。这将降低增强性能并影响网络训练的收敛性，**2)** 类似 U-Net 的网络中使用的跳层连接可能会在最终结果中引入噪声和多余的特征。如何有效地过滤噪音并整合低层次和高层次的特征应该被仔细考虑，**3)** 尽管一些设计和组件被提出来用于 LLIE，但它们大多数是借用或修改自相关的底层视觉任务。在设计网络结构时应考虑低光数据的特点。

3.2 深度模型和 Retinex 理论的结合

如图3(c) 所示，几乎有 1/3 的方法是将深度网络的设计与 Retinex 理论相结合的，例如，设计不同的子网络来估算 Retinex 模型的成分和估算照明图来指导网络的学习。尽管这样的组合可以连接基于深度学习的方法和基于模型的方法，但它们各自的弱点也可能会被引入到最终的模型中：**1)** 基于 Retinex 的 LLIE 方法中使用的理想假设，即反射率是最终的增强结果，仍然会影响最终的结果；**2)** 尽管使用 Retinex 理论，深度网络中过拟合的风险仍然存在。当研究人员将深度学习与 Retinex 理论结合起来时，应该仔细考虑如何取其精华、去其糟粕。

3.3 数据格式

如图3(d) 所示，RGB 数据格式在大多数方法中占主导地位，因为它通常是由智能手机相机、Go-Pro 相机和无人机相机产生的最终图像形式。虽然 raw 格式数据仅限于特殊的传感器，如基于拜尔阵列的传感器，但数据涵盖了更广泛的色域和更高的动态范围。因此，在 raw 数据上训练的深度模型通常可以恢复清晰的细节和高对比度，获得生动的色彩，减少噪音和伪影的影响，并改善极端低照度图像的亮度。在未来对于 LLIE 的研究中，从不同模式的 raw 数据到 RGB 格式的平滑转换，将有可能结合 RGB 数据的便利性和 raw 数据的高质量增强的优势。

3.4 损失函数

在图3(e) 中，LLIE 模型中常用的损失函数包括重建损失 (L_1, L_2, SSIM)、感知损失和平滑度损失。此外，根据不同的需求和表述，还采用了颜色损失、曝光损失、对抗损失等。我们对代表性的损失函数详细说明如下。

重建损失. 不同的重建损失有其优点和缺点。L_2 损失倾向于惩罚较大的错误，但对小错误有一定的容忍度。L_1 损失很好地保护了颜色和亮度，因为无论局部结构如何，错误的权重是相等的。SSIM 损失很好地保留了结构和纹理。详细分析请参考这篇文献 [67]。

感知损失. 感知损失 [68]，特别是特征重构损失，被提出来以约束在特征空间中与 ground truth 相似的结果。该损失可以提高结果的视觉质量。它被定义为增强的结果的特征表示与相应的 ground truth 之间的欧氏距离。特征表示通常是从 ImageNet 数据集 [69] 上预训练的 VGG 网络 [70] 中提取的。

平滑度损失. 为了消除增强结果中的噪声或保持相邻像素的关系，通常使用平滑度损失（TV loss）来约束增强的结果或估算的照度图。

对抗损失. 为了鼓励增强的结果与参考图像没有区别，对抗性学习解决了一个最大-最小的优化问题 [71], [72]。

曝光损失. 作为关键的非参考损失之一，曝光损失衡量在没有非配对参考图像或配对参考图像下增强结果的曝光水平。

在 LLIE 网络中常用的损失函数也可以被用于图像重建网络来实现图像超分辨率 [73]、图像去噪 [74]、图像去抑制 [75], [76], [77] 和图像去模糊 [78]。与这些通用的损失函数不同，专门为 LLIE 设计的曝光损失启发了非参考损失的设计。非参考损失使模型具有更好的泛化能力。考虑图像特性来设计损失函数仍然是一项正在持续进行的研究。

3.5 训练集

图3(f) 报告了各种配对训练数据集在训练弱光增强网络方面的使用情况。这些数据集包括真实世界获取的数据集和合成数据集。我们在表2中列出了它们。

Gamma 校正模拟. 由于其非线性和简单性，Gamma 校正被用来调整视频或静止图像系统中的亮度或三刺激值。它是由一个幂律表达式定义的：

$$V_{\text{out}} = AV_{\text{in}}^{\gamma}, \tag{3}$$

其中输入 V_{in} 和输出 V_{out} 通常在 [0,1] 的范围内。常数 A 在通常情况下被设定为 1。功率 γ 控制输出的亮度。直观地说，当 γ <1 时，输入被调亮，而当 γ >1 时，输入被调暗。输入可以是图像的三个 RGB 通道或与亮度相关的通道，如 CIELab 色彩空间的 L 通道和 YCbCr 色彩空间的 Y 通道。在使用 Gamma 校正法调整亮度相关通道后，色彩空间中的相应通道将被等比例调整，以避免产生伪影和色彩偏差。

为了模拟在真实世界低光场景中拍摄的图像，高斯噪声、泊松噪声或现实噪声被添加到 Gamma 校正后的图像中。使用 Gamma 校正合成的低照度图像可以表示为：

$$I_{\text{low}} = n(g(I_{\text{in}}; \gamma)), \tag{4}$$

其中 n 代表噪声模型，$g(I_{\text{in}}; \gamma)$ 代表 Gamma 值为 γ 的 Gamma 校正函数，I_{in} 是正常光线和高质量图像或与亮度有关的通道。尽管这个函数通过改变 Gamma 值 γ 来产生不同照明水平的低照度图像，由于非线性调整，它往往会在合成的低照度图像中引入伪影和颜色偏差。

随机光照模拟. 根据 Retinex 模型，图像可以被分解为反射分量和光照分量。假设图像内容与光照分量无关，且光照分量

表 2
配对训练数据集的摘要。"Syn " 代表合成图像。

Name	Number	Format	Real/Syn	Video
Gamma Correction	$+\infty$	RGB	Syn	
Random Illumination	$+\infty$	RGB	Syn	
LOL [14]	500	RGB	Real	
SCIE [16]	4,413	RGB	Real	
VE-LOL-L [55]	2,500	RGB	Real+Syn	
MIT-Adobe FiveK [79]	5,000	raw	Real	
SID [14]	5,094	raw	Real	
DRV [18]	202	raw	Real	✓
SMOID [19]	179	raw	Real	✓

表 3
测试数据集摘要.

Name	Number	Format	Application	Video
LIME [5]	10	RGB		
NPE [3]	84	RGB		
MEF [82]	17	RGB		
DICM [83]	64	RGB		
VV[2]	24	RGB		
BBD-100K [84]	10,000	RGB	✓	✓
ExDARK [81]	7,363	RGB	✓	
DARK FACE [80]	6,000	RGB	✓	
VE-LOL-H [55]	10,940	RGB	✓	

中的局部区域具有相同的强度，则可以通过以下方式获得低照度图像:

$$I_{\text{low}} = I_{\text{in}} L, \tag{5}$$

其中 L 是 [0,1] 范围内的一个随机照度值。噪声可以被添加到合成图像中。这样的线性函数可以避免伪影，但强假设要求只能在局部区域具有相同亮度的图像分块上进行合成操作。由于忽略了上下文信息，在这种图像分块上训练的深度模型可能会导致次优的性能。

LOL. LOL [14] 是第一个在真实场景中拍摄的低光/常光图像的配对图像数据集。低光图像是通过改变曝光时间和 ISO 来收集的。LOL 包含 500 对大小为 400×600 的低光/常光图像，以 RGB 格式保存。

SCIE. SCIE 是一个由低对比度和良好对比度图像对组成的多曝光配对图像数据集。它包括 589 个室内和室外场景的多重曝光图像序列。每个序列有 3 到 18 张不同曝光度的低对比度图像，因此总共包含 4413 张多重曝光图像。这 589 张高质量的参考图像是通过从 13 种有代表性的增强算法的结果中挑选出来的。也就是说，许多多重曝光的图像都有相同的高对比度参考图像。这些图像的分辨率是 3,000×2,000 或是 6,000×4,000。SCIE 中的图像是以 RGB 格式保存的。

MIT-Adobe FiveK. MIT-Adobe FiveK [79] 被收集用于全局色调调整，但也已被用于 LLIE。这是因为输入的图像具有低亮度和低对比度。MIT-Adobe FiveK 包含 5,000 张图片，每张图片都由 5 位专业的摄影师进行修图，以达到视觉上的令人满意的效果，类似于明信片。这些图像都是 raw 格式的。为了训练能够处理 RGB 格式图像的网络，人们需要使用 Adobe Lightroom 对图像进行预处理，并按照这个程序[1]将其保存为 RGB 格式。图像通常被调整为长边为 500 像素的分辨率。

SID. SID [12] 包含 5,094 张 raw 格式短曝光图像，每张都有相应的长曝光参考图像。长曝光图像的不同参考图像的数量是 424 张，换句话说，多个短曝光图像对应于同一个长曝光参考图像。这些图像是用两台相机拍摄的：索尼 α7S II 和富士 X-T2 在室内和室外场景中拍摄。因此，图像具有不同的传感器模式（索尼相机的拜尔传感器和富士相机的 APS-C X-Trans 传感器）。索尼的分辨率为 4,240×2,832，富士的分辨率为 6,000×4,000。通常，长曝光的图像由 libraw（一个 raw 图像处理库）处理并保存在 RGB 色彩空间，并随机裁剪 512×512 的分块用于训练。

VE-LOL. VE-LOL [55] 由两个子集组成：配对的 VE-LOL-L，用于训练和评估 LLIE 方法；非配对的 VE-LOL-H，用于评估 LLIE 方法对人脸检测的影响。具体来说，VE-LOL-L 包括 2,500 组成对的图像。其中，1,000 对是合成的，而 1,500 对是真实拍摄的。VE-LOL-H 包括 10,940 张不成对的图像，其中人脸是用边界框手动标注的。

DRV. DRV [18] 包含 202 个静态的 raw 视频，每个视频都有一个相应的长曝光的 ground truth。每段视频都是在连续拍摄模式下以每秒约 16 至 18 帧的速度拍摄的，最多有 110 帧。图像是由索尼 RX100 VI 相机在室内和室外场景中拍摄的，因此都是拜尔模式的 raw 格式。分辨率为 3,672×5,496。

SMOID. SMOID [19] 包含 179 组由共轴光学系统拍摄的配对视频，每个视频有 200 帧。因此，SMOID 包括 35,800 个拜尔列阵的极低光照 raw 数据及其相应的光照良好的 RGB 对照数据。SMOID 由不同光照条件下的移动车辆和行人组成。

上述的配对训练数据集仍面临以下问题的挑战：**1)** 在合成数据上训练的深层模型在处理真实世界的图像和视频时，由于合成数据和真实数据之间的差距，可能会引入伪影和颜色偏差。**2)** 真实训练数据的规模和多样性并不令人满意，因此一些方法采用合成数据来增加训练数据。这可能会导致次优的增强，并且 **3)** 由于运动、硬件和环境的影响，输入的图像和相应的 ground truths 可能存在错位。这将影响使用像素级损失函数训练的深度网络的性能。

3.6 测试数据集

除了配对数据集中的测试子集外 [12], [14], [16], [18], [19], [55], [79]，还有一些从相关作品中收集的测试数据或常用于实验比较的数据。此外，一些数据集，如黑暗中的人脸检测 [80] 和低照度图像中的检测和识别 [81]，被用来测试 LLIE 对高层视觉任务的影响。我们在表3中总结了常用的测试数据集，并介绍了以下有代表性的测试数据集。

BBD-100K. BBD-100K [84] 是最大的驾驶视频数据集，其中有 10,000 个视频，拍摄于 1,100 小时的驾驶过程，涉及一天中许多不同的时间、天气条件和驾驶场景，以及 10 个任务

1. https://github.com/nothinglo/Deep-Photo-Enhancer/issues/38#issuecomment-449786636

2. https://sites.google.com/site/vonikakis/datasets

注释。BBD-100K 中夜间拍摄的视频被用来验证 LLIE 对高级视觉任务的影响以及在真实场景中的增强性能。

ExDARK. ExDARK [81] 数据集是为低照度图像中的物体检测和识别而建立的。ExDARK 数据集包含了 7,363 张从极低光环境到黄昏的低光图像，其中有 12 个物体类别，并附有图像类别标签和局部物体边界框的标注。

DARK FACE. DARK FACE [80] 数据集包含 6,000 张在夜间拍摄的低照度图像，每张图像都标注了人脸的边界框。

从图3 (g) 和表1中，我们可以看到，人们更喜欢在实验中使用自己收集的测试数据。主要原因有三个方面。**1)** 除了配对数据集的测试子集，没有公认的评估基准，**2)** 常用的测试集存在一些缺陷，如规模小（有些测试集只包含 10 张图像），内容和光照属性重复，以及未知的实验设置，以及 **3)** 一些常用的测试数据并非最初为评估 LLIE 而收集。一般来说，目前的测试数据集可能会导致偏袒和不公平的比较。

3.7 评价指标

除了基于人类感知的主观评价，图像质量评估（IQA）指标，包括全参考和非参考 IQA 指标，都能够客观地评价图像质量。此外，用户调研、可训练参数的数量、FLOPs、运行时间和应用也反映了 LLIE 模型的性能，如图3（h）所示。我们将详细介绍如下。

PSNR 和 MSE. PSNR 和 MSE 是广泛使用的 IQA 指标。它们总是非负值，接近无限（PSNR）和零（MSE）的值更好。然而，像素级的 PSNR 和 MSE 可能会提供一个不准确的图像质量视觉感知的指示，因为它们忽略了相邻像素的关系。

MAE. MAE 代表平均绝对误差，作为成对的观测值之间误差的衡量标准。两者的 MAE 值越小，相似度就越高。

SSIM. SSIM 被用来衡量两幅图像之间的相似性。它是一个基于感知的模型，将图像退化视为结构信息的感知变化。其值为 1 只有在两组完全相同的数据情况下才能达到，表明结构完全相似。

LOE. LOE 代表反映增强图像自然度的亮度顺序误差。对于 LOE 来说，LOE 值越小，亮度顺序的保存的就越好。

应用. 除了提高视觉质量，图像增强的目的之一是为高级视觉任务服务。因此，LLIE 针对高层次视觉应用的影响通常被研究，以验证不同方法的性能。目前在 LLIE 中使用的评价方式需要在几个方面加以改进:

1) 尽管 PSNR、MSE、MAE 和 SSIM 是经典而流行的指标，但它们仍然远远不能捕捉到人类的真实视觉感受。**2)** 有些指标最初并不是为低光图像设计的。它们是用来评估图像信息和对比度的保真度的。使用这些指标可以反映出图像的质量，但它们与低照度增强的真正目的相去甚远。**3)** 除了 LOE 指标外，还缺乏专门为低光图像设计的指标。此外，也没有评估低照度视频增强的指标，并且 **4)** 我们希望有一个能够平衡人类视觉和机器感知的指标。

表 4
LLIV-Phone 数据集摘要。LLIV-Phone 数据集包含由 18 个不同的手机摄像头拍摄的 120 个视频（45,148 张图片）。“#Video” 和 “#Image” 分别代表视频和图像的数量。

Phone's Brand	#Video	#Image	Resolution
iPhone 6s	4	1,029	1920×1080
iPhone 7	13	6,081	1920×1080
iPhone7 Plus	2	900	1920×1080
iPhone8 Plus	1	489	1280×720
iPhone 11	7	2,200	1920×1080
iPhone 11 Pro	17	7,739	1920×1080
iPhone XS	11	2,470	1920×1080
iPhone XR	16	4,997	1920×1080
iPhone SE	1	455	1920×1080
Xiaomi Mi 9	2	1,145	1920×1080
Xiaomi Mi Mix 3	6	2,972	1920×1080
Pixel 3	4	1,311	1920×1080
Pixel 4	3	1,923	1920×1080
Oppo R17	6	2,126	1920×1080
Vivo Nex	12	4,097	1280×720
LG M322	2	761	1920×1080
OnePlus 5T	1	293	1920×1080
Huawei Mate 20 Pro	12	4,160	1920×1080

4 基准测试与实证分析

本节提供实证分析，并强调了基于深度学习的 LLIE 的一些关键挑战。为了便于分析，我们提出了一个低照度图像和视频数据集，以检验不同解决方案的性能。我们还开发了第一个在线平台，LLIE 模型的结果可以通过一个用户友好的网页界面产生。在本节中，我们对几个基准和我们提出的数据集进行了广泛的评估。

在实验中，我们比较了 13 种有代表性的基于 RGB 格式的方法，包括 8 种基于监督学习的方法（LLNet [11], LightenNet [15], Retinex-Net [14], MBLLEN [13], KinD [21], KinD++ [61], TBEFN [30], DSLR [31]），一种基于无监督学习的方法（EnlightenGAN [36]），一种基于半监督学习的方法（DRBN [43]），以及三种基于零次学习的方法（ExCNet [37], Zero-DCE [38], RRDNet [39]）。另一方面，我们还比较了两种基于 raw 格式的方法，包括 SID [85] 和 EEMEFN [26]。请注意，基于 RGB 格式的方法在 LLIE 中占大多数。此外，大多数基于 raw 格式的方法都没有公布其代码。因此，我们选择了两种有代表性的方法来提供实验分析和见解。对于所有被比较的方法，我们使用公开的代码来产生它们的结果，以便进行公平的比较。

4.1 新低照度图像和视频数据集

我们提出了一个低照度图像和视频数据集，称为 LLIV-Phone，以全面彻底地验证 LLIE 方法的性能。LLIV-Phone 是同类中最大和最具挑战性的真实世界测试数据集。特别是，该数据集包含 120 个视频（45,148 张图片），由 18 个不同的手机摄像头拍摄，包括 iPhone 6s、iPhone 7、iPhone 7 Plus、iPhone 8 Plus、iPhone 11、iPhone 11 Pro、iPhone XS、iPhone XR、iPhone SE、小米 9、小米 MIX 3、Pixel 3、Pixel 4、Oppo R17、Vivo Nex、LG M322、OnePlus 5T、华为 Mate 20 Pro，在不同照明条件（例如弱光、曝光不足、月光、黄昏、黑暗、极暗、

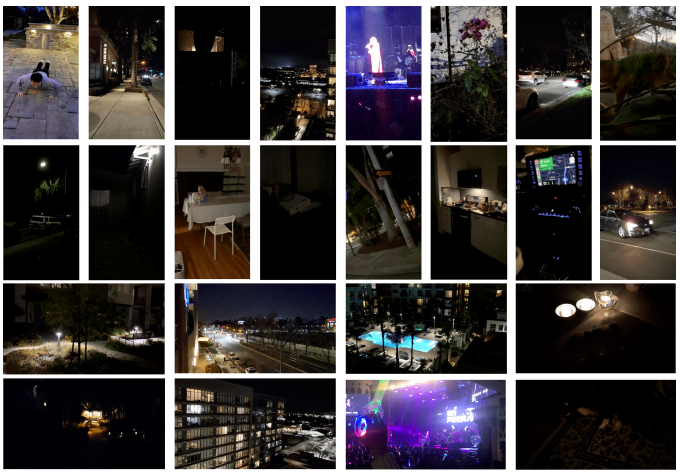


图 4. 从提出的 LLIV-Phone 数据集中部分图像样张。这些图像和视频是由不同的设备在不同的照明条件和场景下拍摄的。

背光、非均匀光照和彩色光照）的室内和室外场景。表4中提供了 LLIV-Phone 数据集的摘要。

我们在图4中展示了 LLIV-Phone 数据集的几个样本。LLIV-Phone 的数据集可在项目页面获得。

这个具有挑战性的数据集是在真实场景中收集的，包含各种低照度图像和视频。因此，它适用于评估不同的低光图像和视频增强模型的泛化能力。值得一提的是，该数据集可以作为无监督学习的训练数据集和合成方法的参考数据集，以生成真实的低照度数据。

4.2 在线评估平台

不同的深度模型可以在不同的平台上实现，如 Caffe、Theano、TensorFlow 和 PyTorch。因此，不同的算法需要不同的配置、GPU 版本和硬件规格。这样的要求让许多研究人员望而却步，特别是对于刚进入这个领域的初学者，他们甚至可能没有 GPU 资源。为了解决这些问题，我们开发了一个 LLIE 在线平台，称为 LLIE-Platform，可在 http://mc.nankai.edu.cn/ll/ 访问。

截至本报告提交之日，LLIE-Platform 涵盖了 14 种流行的基于深度学习的 LLIE 方法，包括 LLNet [11]、LightenNet [15]、Retinex-Net [14]、Enlighten-GAN [36]、MBLLEN [13]、KinD [21]。KinD++ [61]、TBEFN [30]、DSLR [31]、DRBN [43]、ExCNet [37]、Zero-DCE [38]、Zero-DCE++ [40] 和 RRDNet [39]，其中任何输入的结果都可以通过一个用户友好的网页界面产生。我们将定期在这个平台上提供更新新的方法。我们希望这个 LLIE 平台能够服务于不断增长的研究界，为用户提供一个灵活的界面来运行现有的基于深度学习的 LLIE 方法并开发他们自己的新 LLIE 方法。

4.3 基准测试结果

为了定性和定量地评估不同的方法，除了提议的 LLIV-Phone 数据集，我们还采用了常用的 LOL [14] 和 MIT-Adobe FiveK [79] 数据集，用于测试基于 RGB 格式的方法，以及 SID [85] 数据集用于基于 raw 格式的方法。更多的视觉结果可以在补充材料中找到。由不同手机摄像头拍摄的真实低光视频的比较结果可以在 YouTube: https://www.youtube.com/watch?v=Elo9TkrG5Oo&t=6s上找到。

我们从 LLIV-Phone 数据集的每个视频中平均选取 5 张图片，组成一个总共有 600 张图片的图片测试数据集（记为 LLIV-Phone-imgT）。此外，我们从 LLIV-Phone 数据集的每个手机品牌的视频中随机选择一个视频，形成一个总共有 18 个视频的视频测试数据集（记为 LLIV-Phone-vidT）。我们将 LLIV-Phone-imgT 和 LLIV-Phone-vidT 中的帧的分辨率减半，因为一些基于深度学习的方法不能处理全分辨率的测试图像和视频。对于 LOL 数据集，我们采用原始的测试集，包括 15 张在真实场景中拍摄的低照度图像进行测试，表示为 LOL-test。对于 MIT-Adobe FiveK 数据集，我们遵循 Chen 等人 [47] 的提议，将图像解码为 PNG 格式，并使用 Lightroom 将其调整成长边为 512 像素的图片。我们采用与 Chen 等人 [47] 相同的测试数据集，即 MIT-Adobe FiveK-test，包括 500 张由专业人士修图过的图像作为相应的 ground truths。对于 SID 数据集，我们使用 EEMEFN [26] 中默认的测试集进行公平的比较，表示为 SID-test（SID-test-Bayer 和 SID-test-X-Trans），它是 SID [85] 的一个部分测试集。SID-test-Bayer 包括 93 张拜耳列阵的图像，而 SID-test-X-Trans 包括 94 张 APS-C X-Trans 模式的图像。

定性比较. 我们首先在图5和图6中展示了不同方法在 LOL-test 和 MIT-Adobe FiveK-test 数据集样本中生成的结果。

如图5所示，所有方法都提高了输入图像的亮度和对比度。然而，当生成的结果与 ground truth 比较时，它们都没有成功准确地恢复输入图像的颜色。特别是，LLNet [11] 产生了模糊的结果。LightenNet [15] 和 RRDNet [39] 产生了曝光不足的结果，而 MBLLEN [13] 和 ExCNet [37] 则使图像曝光过度。KinD [21]、KinD++ [61]、TBEFN [30]、DSLR [31]、EnlightenGAN [36] 和 DRBN [43] 引入了明显的伪影。在图6中，LLNet [15]、KinD++ [61]、TBEFN [30] 和 RRDNet [39] 产生过度曝光的结果。Retinex-Net [14]、KinD++ [61] 和 RRDNet [39] 在结果中产生了伪影和模糊现象。

我们发现，MIT-Adobe FiveK 数据集的 ground truths 仍然包含一些黑暗区域。这是因为该数据集最初是为全局图像修饰而设计的，而恢复低光区域不是其主要的任务。我们还观察到，LOL 数据集和 MIT-Adobe FiveK 数据集的输入图像的噪声相对少，这与真实的低光场景不同。尽管有些方法 [28], [31], [60] 将 MIT-Adobe FiveK 数据集作为训练或测试数据集，但我们认为该数据集不适合 LLIE 的任务，因为它的 ground truth 不适合做 LLIE 的 ground truth。

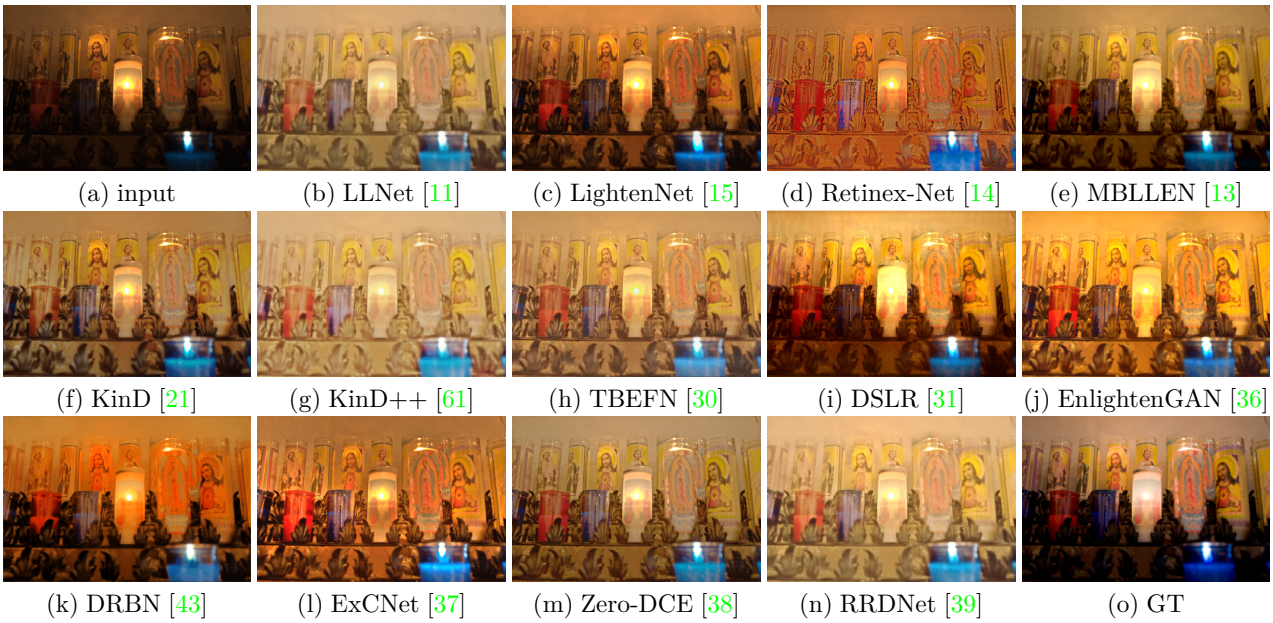
为了考察不同方法的泛化能力，我们对 LLIV-Phine-imgT 数据集中的图像进行了比较。不同方法的视觉结果显示在图7和图8 中。如图7所示，所有的方法都不能有效地对低照度图像提高亮度和去除噪声。此外，Retinex-Net [14]、

(a) input (b) LLNet [11] (c) LightenNet [15] (d) Retinex-Net [14] (e) MBLLEN [13]

(f) KinD [21] (g) KinD++ [61] (h) TBEFN [30] (i) DSLR [31] (j) EnlightenGAN [36]

(k) DRBN [43] (l) ExCNet [37] (m) Zero-DCE [38] (n) RRDNet [39] (o) GT

图 5. 不同方法在 LOL 测试数据集低照度图像样张上的视觉结果。



(a) input (b) LLNet [11] (c) LightenNet [15] (d) Retinex-Net [14] (e) MBLLEN [13]

(f) KinD [21] (g) KinD++ [61] (h) TBEFN [30] (i) DSLR [31] (j) EnlightenGAN [36]

(k) DRBN [43] (l) ExCNet [37] (m) Zero-DCE [38] (n) RRDNet [39] (o) GT

图 6. 不同方法在 MIT-Adobe FiveK 测试数据集低照度图像样张上的视觉结果。

MBLLEN [13] 和 DRBN [43] 产生了明显的伪影。在图8中，所有的方法都提高了该输入图像的亮度。然而，只有 MBLLEN [13] 和 RRDNet [39] 获得了视觉上很好的增强，没有颜色偏差、伪影和过曝/欠曝。值得注意的是，对于有光源的区域，没有一种方法可以在不放大这些区域周围的噪声的情况下提高图像的亮度。将光源纳入 LLIE 的考虑范围将是一个有趣的探索方向。结果表明，增强 LLIV-Phine-imgT 数据集的图像是很困难的。由于现有的 LLIE 方法的泛化能力有限，真实的低光图像让大多数 LLIE 方法失效。潜在的原因是使用合成训练数据、小比例的训练数据或不现实的假设，如局部光照一致性和在这些方法中把反射成分作为 Retinex 模型的最终结果。

我们在图9中进一步展示了基于 raw 格式的方法的视觉比较。如图所示，输入的原始数据有明显的噪音。SID [12] 和 EEMEFN [26] 都可以有效地消除噪声的影响。与 SID 中使用的简单的 U-Net 结构相比，EEMEFN 的更复杂的结构获得了更好的亮度恢复。然而，他们的结果与相应的 GT 相差甚远，特别是对于 APS-C X-Trans 模式的输入。

定量比较. 对于带有 ground truth 的测试集，即 LOL-test、MIT-Adobe FiveK-test 和 SID-test，我们采用 MSE、PSNR、SSIM [86] 和 LPIPS [87] 指标来定量比较不同的方法。LPIPS 是一个基于深度学习的图像质量评估指标，它通过深度视觉

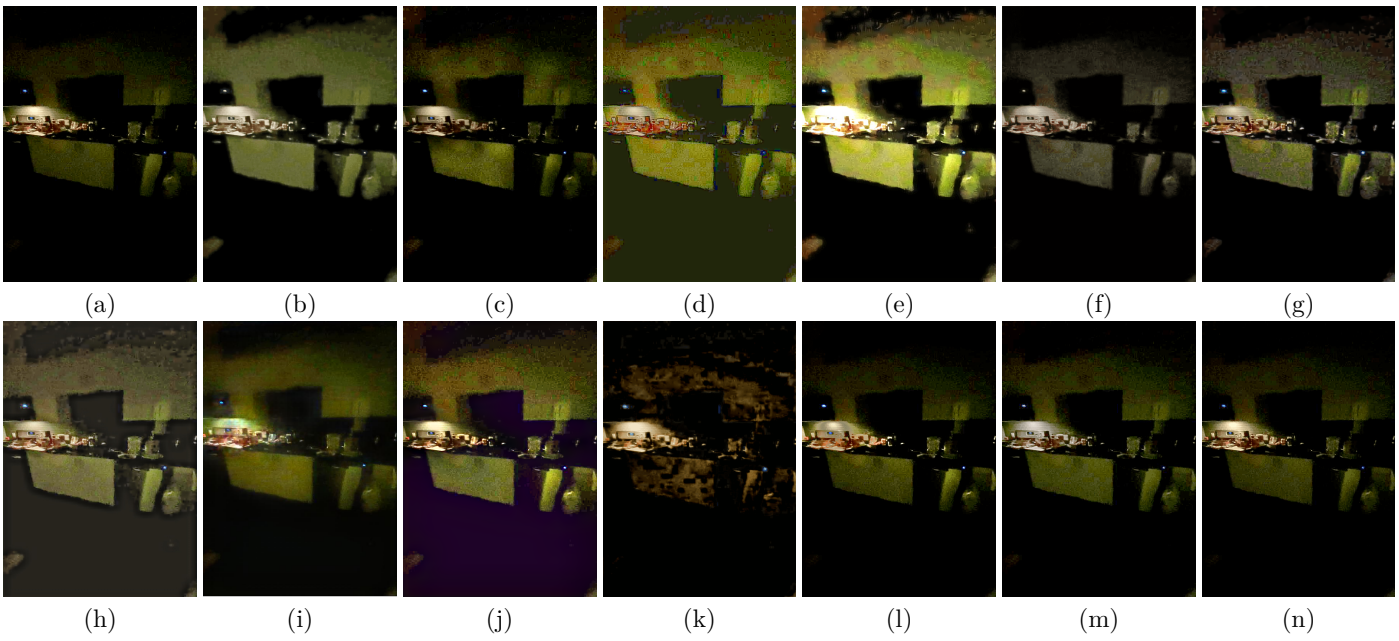


图 7. 不同方法在 LLIV-Phon-imgT 数据集取样的低照度图像上的视觉效果。(a) input. (b) LLNet [11]. (c) LightenNet [15]. (d) Retinex-Net [14]. (e) MBLLEN [13]. (f) KinD [21]. (g) KinD++ [61]. (h) TBEFN [30]. (i) DSLR [31]. (j) EnlightenGAN [36]. (k) DRBN [43]. (l) ExCNet [37]. (m) Zero-DCE [38]. (n) RRDNet [39].

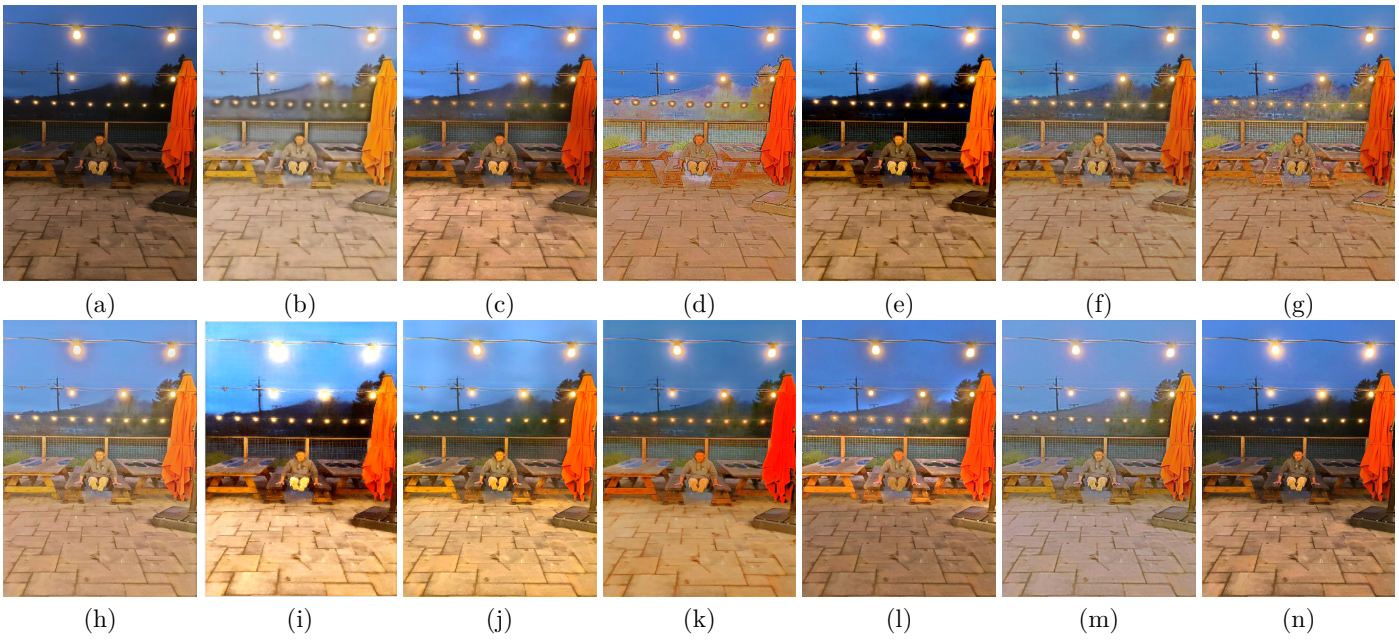
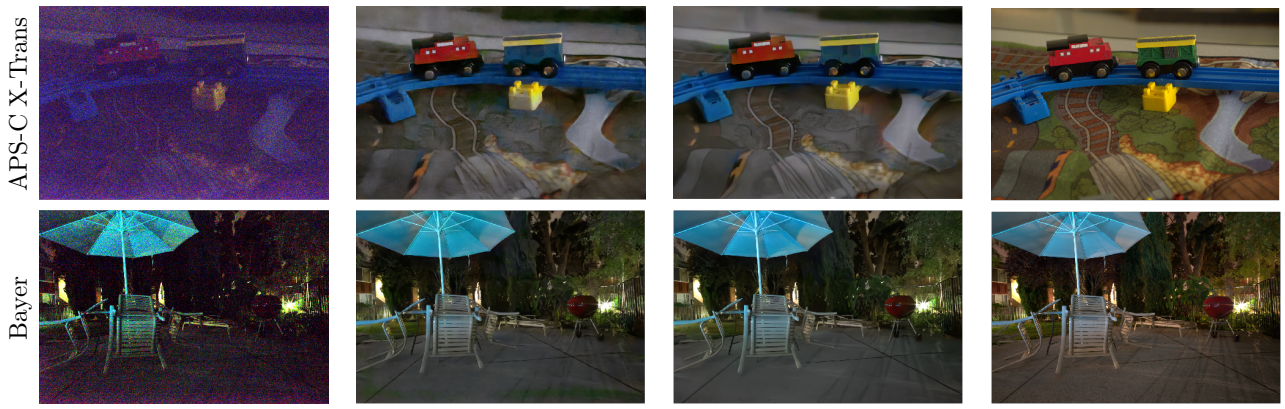


图 8. 不同方法在 LLIV-Phon-imgT 数据集取样的低照度图像上的视觉效果。(a) input. (b) LLNet [11]. (c) LightenNet [15]. (d) Retinex-Net [14]. (e) MBLLEN [13]. (f) KinD [21]. (g) KinD++ [61]. (h) TBEFN [30]. (i) DSLR [31]. (j) EnlightenGAN [36]. (k) DRBN [43]. (l) ExCNet [37]. (m) Zero-DCE [38]. (n) RRDNet [39].

表征来衡量一个结果和其对应的 ground truth 之间的感知相似性。对于 LPIPS，我们采用基于 AlexNet 的模型来计算感知相似度。LPIPS 值越低，说明结果在感知相似度方面越接近于其对应的 ground truth。在表5和表6中，我们分别显示了基于 RGB 格式方法和基于 raw 格式方法的定量结果对比。

如表5所示，在 LOL-test 和 MIT-Adobe FiveK-test 数据集上，基于监督学习的方法的得分优于基于无监督学习、半监督学习和零点学习的方法。其中，LLNet [11] 在 LOL-test 数据集上获得了最好的 MSE 和 PSNR 值；然而，它在 MIT-Adobe FiveK-test 数据集上的性能有所下降。这可能是由于 LLNet [11] 对 LOL 数据集的偏向造成的，因为它是用 LOL 训练数据集训练的。对于 LOL-test 数据集，TBEFN [30] 获得了最高的 SSIM 值，而 KinD [21] 则获得了最低的 LPIPS 值。尽管有些方法是在 LOL 训练数据集上训练出来的，但在 LOL-test 数据集上，没有任何一个算法在这四种评价指标上明显胜出。对于 MIT-Adobe FiveK-test 数据集，尽管 MBLLEN [13] 是



(a) inputs (b) SID [85] (c) EEMEFN [26] (d) GT

图 9. 从 SID-test-Bayer 和 SID-test-X-Trans 测试数据集取样的两幅 raw 格式低照度图像上不同方法的视觉结果。输入图像被增强以获得更佳可视效果。

表 5
对 LOL-test 和 MIT-Adobe FiveK-test 数据集的 MSE($\times 10^3$)，PSNR (in dB)，SSIM [86]，和 LPIPS [87] 进行定量比较。最佳结果用红色表示，而第二和第三好的结果分别用蓝色和紫色表示。

Learning	Method	LOL-test				MIT-Adobe FiveK-test			
		MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓
	input	12.613	7.773	0.181	0.560	1.670	17.824	0.779	0.148
SL	LLNet [11]	1.290	17.959	0.713	0.360	4.465	12.177	0.645	0.292
	LightenNet [15]	7.614	10.301	0.402	0.394	4.127	13.579	0.744	0.166
	Retinex-Net [14]	1.651	16.774	0.462	0.474	4.406	12.310	0.671	0.239
	MBLLEN [13]	1.444	17.902	0.715	0.247	1.296	19.781	0.825	0.108
	KinD [21]	1.431	17.648	0.779	0.175	2.675	14.535	0.741	0.177
	KinD++ [61]	1.298	17.752	0.760	0.198	7.582	9.732	0.568	0.336
	TBEFN [30]	1.764	17.351	0.786	0.210	3.865	12.769	0.704	0.178
	DSLR [31]	3.536	15.050	0.597	0.337	1.925	16.632	0.782	0.167
UL	EnlightenGAN [36]	1.998	17.483	0.677	0.322	3.628	13.260	0.745	0.170
SSL	DRBN [43]	2.359	15.125	0.472	0.316	3.314	13.355	0.378	0.281
ZSL	ExCNet [37]	2.292	15.783	0.515	0.373	2.927	13.978	0.710	0.187
	Zero-DCE [38]	3.282	14.861	0.589	0.335	3.476	13.199	0.709	0.203
	RRDNet [39]	6.313	11.392	0.468	0.361	7.057	10.135	0.620	0.303

表 6
对 SID-test 数据集的 MSE($\times 10^3$)，PSNR (in dB)，SSIM [86]，和 LPIPS [87] 进行定量比较。最佳结果用红色表示。为了计算输入 raw 数据的量化分数，我们使用 Chen 等人 [85] 提供的相应的相机 ISP 管道，将 raw 数据转为 RGB 格式。

Learning	Method	SID-test–Bayer				SID-test–X-Trans			
		MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓
	input	5.378	11.840	0.063	0.711	4.803	11.880	0.075	0.796
SL	SID [85]	0.140	28.614	0.757	0.465	0.235	26.663	0.680	0.586
	EEMEFN [26]	0.126	29.212	0.768	0.448	0.191	27.423	0.695	0.546

在合成训练数据上训练的，但在四个评价指标上都优于其他方法。尽管如此，MBLLEN [13] 仍然不能同时在这两个测试数据集上获得最佳性能。

如表6所示，SID [85] 和 EEMEFN [26] 都提高了 raw 输入数据的质量。与 SID 的分数相比，EEMEFN 在不同的原始数据模式和评价指标上都取得了一贯的更佳性能。

对于 LLIV-Phone-imgT 测试集，我们使用非参考 IQA 指标，即 NIQE [88]、感知指数（PI）[88], [89], [90]、LOE [3] 和 SPAQ [91] 来定量比较不同方法。就 LOE 而言，LOE 值越小，明亮度顺序保留的就越好。就 NIQE 而言，NIQE 值越小，视觉质量就越好。PI 值越低，表示感知质量越好。SPAQ 是为智能手机摄影的感观质量评估而设计的。SPAQ 值越大，说明智能手机摄影的感观质量越好。具体结果见表7。

观察表7，我们可以发现 Retinex-Net [14]、KinD++ [61] 和 EnlightenGAN [36] 的性能相对优于其他方法。Retinex-Net [14] 获得了最好的 PI 和 SPAQ 分数。这些分数表明 Retinex-Net [14] 所增强的结果具有良好的感知质量。然而，从图 7(d) 和图 8(d) 来看，Retinex-Net [14] 的结果显然受到伪影和颜色偏差的困扰。此外，KinD++ [61] 获得了最低的 NIQE 分数，而原始输入获得了最低的 LOE 分数。对于现有的标准 LOE 指标，我们质疑亮度顺序是否能有效反映增强性能。总的来说，在某些情况下，非参考 IQA 指标在评估增强的低照度图像的质量时出现了偏差。

为了准备 LLIV-vidT 测试集的视频，我们首先放弃了连

表 7
LLIV-Phine-imgT 数据集在 NIQE [88], LOE [3], PI [88], [89], [90], 和 SPAQ [91] 方面的定量比较。最佳结果用红色表示，而第二和第三好的结果分别用蓝色和紫色表示。

Learning	Method	LoLi-Phone-imgT			
		NIQE↓	LOE ↓	PI↓	SPAQ↑
	input	6.99	0.00	5.86	44.45
SL	LLNet [11]	5.86	5.86	5.66	40.56
	LightenNet [15]	5.34	952.33	4.58	45.74
	Retinex-Net [14]	5.01	790.21	3.48	50.95
	MBLLEN [13]	5.08	220.63	4.27	42.50
	KinD [21]	4.97	405.88	4.37	44.79
	KinD++ [61]	4.73	681.97	3.99	46.89
	TBEFN [30]	4.81	552.91	4.30	44.14
	DSLR [31]	4.77	447.98	4.31	41.08
UL	EnlightenGAN [36]	4.79	821.87	4.19	45.48
SSL	DRBN [43]	5.80	885.75	5.54	42.74
ZSL	ExCNet [37]	5.55	723.56	4.38	46.74
	Zero-DCE [38]	5.82	307.09	4.76	46.85
	RRDNet [39]	5.97	142.89	4.84	45.31

表 8
对 LLIV-Phone-vidT 数据集的平均亮度差异（ALV）得分进行定量比较。最佳结果用红色表示，而第二和第三好的结果分别用蓝色和紫色表示。

Learning	Method	LoLi-Phone-vidT
		ALV↓
	input	185.60
SL	LLNet [11]	85.72
	LightenNet [15]	643.93
	Retinex-Net [14]	94.05
	MBLLEN [13]	113.18
	KinD [21]	98.05
	KinD++ [61]	115.21
	TBEFN [30]	58.69
	DSLR [31]	175.35
UL	EnlightenGAN [36]	90.69
SSL	DRBN [43]	115.04
ZSL	ExCNet [37]	1375.29
	Zero-DCE [38]	117.22
	RRDNet [39]	147.11

续帧中没有明显物体的视频，于是一共选择了 10 个视频。对于每个视频，我们选择一个出现在所有帧中的物体。然后我们使用一个追踪器 [92] 来追踪输入视频连续帧中的物体，并确保同一物体被标在边界框中。我们丢弃对物体追踪不准确的帧然后收集每一帧中边界框的坐标。我们采用这些坐标来裁剪通过不同方法增强的结果中的相应区域，并计算物体在连续帧中的平均亮度差异（ALV）得分：$ALV = \frac{1}{N}\sum_{i=1}^{N}(L_i - L_{\text{avg}})^2$，其中 N 是视频的帧数，L_i 代表第 i 帧中边界框区域的平均亮度值，L_{avg} 表示视频中所有边界框区域的平均亮度值。较低的 ALV 值表明增强后的视频具有更好的时间一致性。表8显示了不同方法在 LLIV-vidT 测试集的 10 个视频中的平均 ALV 值。不同方法在每个视频上的 ALV 值可以在补充材料中找到。此外，我们按照 Jiang 和 Zheng [19] 的做法，在补充材料中绘制了他们的亮度曲线。

如表8所示，TBEFN [30] 在 ALV 值方面获得了最好的时间一致性，而 LLNet [11] 和 EnlightenGAN [36] 分别排名第二和第三。相比之下，ExCNet [37] 的 ALV 值作为最差的表现，达到 1375.29。这是因为基于零参考学习的 ExCNet [37] 在增强连续帧方面的表现不稳定。ExCNet [37] 可以有效地提高某些帧的亮度，而对其他帧的效果却不好。

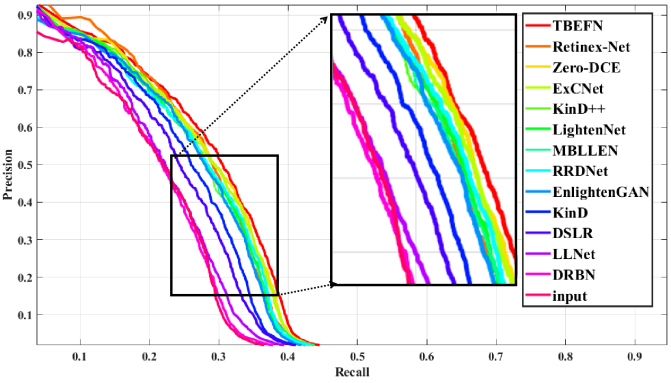


图 10. 黑暗中人脸检测的 P-R 曲线

4.4 计算复杂度

在表9中，我们比较了基于 RGB 格式方法的计算复杂度，包括运行时间、可训练参数和使用 NVIDIA 1080Ti GPU 对 32 幅大小为 1200×900×3 图像的平均 FLOPs。为了进行公平的比较，我们省略了 LightenNet [15]，因为只有其代码的 CPU 版本是公开的。此外，我们没有汇报 ExCNet [37] 和 RRDNet [39] 的 FLOPs，因为这个数字取决于输入图像（不同的输入需要不同的迭代数）。

如表9所示，Zero-DCE [38] 的运行时间最短，因为它只通过一个轻量级网络估算几个曲线的参数。因此，它的可训练参数和 FLOPs 的数量要少得多。此外，LightenNet [15] 的可训练参数数量和 FLOPs 是所有比较方法中最少的。这是因为 LightenNet [15] 通过一个由四个卷积层组成的微小网络来估算输入图像的照度图。相比之下，LLNet [11] 和 KinD++ [61] 的 FLOPs 非常大，分别达到 4124.177G 和 12238.026G。基于零次学习的 ExCNet [37] 和 RRDNet [39] 的运行时间很长，因为优化过程很耗时。

4.5 基于应用的评测

我们研究了低照度图像增强方法在黑暗中人脸检测方面的表现。按照 Guo 等人 [38] 的设定，我们使用了 DARK FACE [80] 数据集，该数据集由在黑暗中拍摄的人脸图像组成。由于测试集的边界框没有公开，我们对从训练集和验证集中随机抽取的 500 张图像进行了评估。在 WIDER FACE [93] 数据集上训练的 Dual Shot Face Detector (DSFD) [94] 被用作面部检测器。我们将不同 LLIE 方法的结果反馈给 DSFD，并在图中描绘了在 0.5 IoU 阈值下的精度-召回（P-R）曲线10。我们使用 DARK FACE 数据集 [80] 中提供的评估工具[3]比较了不同 IoU 阈值下的平均精度（AP），见表10。

如图10所示，所有基于深度学习的解决方案都提高了黑暗中人脸检测的性能，这表明基于深度学习的 LLIE 解决方

3. https://github.com/Ir1d/DARKFACE_eval_tools

表 9
以 runtime（秒）、可训练参数（#Parameters）数量（M）和 FLOPs（G）对计算复杂性进行定量比较。最佳结果用红色表示，而第二和第三好的结果分别用蓝色和紫色表示。'-' 表示该结果未获得。

Learning	Method	RunTime↓	#Parameters ↓	FLOPs↓	Platform
SL	LLNet [11]	36.270	17.908	4124.177	Theano
	LightenNet [15]	-	0.030	30.540	MATLAB
	Retinex-Net [14]	0.120	0.555	587.470	TensorFlow
	MBLLEN [13]	13.995	0.450	301.120	TensorFlow
	KinD [21]	0.148	8.160	574.954	TensorFlow
	KinD++ [61]	1.068	8.275	12238.026	TensorFlow
	TBEFN [30]	0.050	0.486	108.532	TensorFlow
	DSLR [31]	0.074	14.931	96.683	PyTorch
UL	EnlightenGAN [36]	0.008	8.637	273.240	PyTorch
SSL	DRBN [43]	0.878	0.577	196.359	PyTorch
ZSL	ExCNet [37]	23.280	8.274	-	PyTorch
	Zero-DCE [38]	0.003	0.079	84.990	PyTorch
	RRDNet [39]	167.260	0.128	-	PyTorch

表 10
在黑暗中人脸检测的不同 IoU 阈值下的 AP 的比较。最佳结果用红色表示，而第二和第三好的结果分别用蓝色和紫色表示。

Learning	Method	IoU thresholds		
		0.5	0.6	0.7
	input	0.195	0.061	0.007
SL	LLNet [11]	0.208	0.063	0.006
	LightenNet [15]	0.249	0.085	0.010
	Retinex-Net [14]	0.261	0.101	0.013
	MBLLEN [13]	0.249	0.092	0.010
	KinD [21]	0.235	0.081	0.010
	KinD++ [61]	0.251	0.090	0.011
	TBEFN [30]	0.268	0.099	0.011
	DSLR [31]	0.223	0.067	0.007
UL	EnlightenGAN [36]	0.246	0.088	0.011
SSL	DRBN [43]	0.199	0.061	0.007
ZSL	ExCNet [37]	0.256	0.092	0.010
	Zero-DCE [38]	0.259	0.092	0.011
	RRDNet [39]	0.248	0.083	0.010

案在黑暗中人脸检测的有效性。如表10所示，不同 IoU 阈值下最佳方法的 AP 得分在 0.268 到 0.013 之间，而不同 IoU 阈值下的原始输入 AP 得分非常低。这些结果表明，仍有改进的余地。值得注意的是，Retinex-Net [14]、Zero-DCE [38] 和 TBEFN [30] 在黑暗中实现了相对鲁棒的人脸检测性能。我们在图11中展示了不同方法的视觉结果。尽管 Retinex-Net [14] 在 AP 得分上比其他方法表现更好，但它的视觉结果包含明显的人工痕迹和不自然的纹理。总的来说，Zero-DCE [38] 在黑暗中人脸检测的 AP 得分和感知质量之间取得了良好的平衡。请注意，黑暗中人脸检测的结果不仅与增强的结果有关，还与人脸检测器有关，包括检测器模型和检测器的训练数据。这里，我们只以预先训练好的 DSFD [94] 为例，在一定程度上验证了不同方法的低照度图像增强性能。

4.6 讨论

从实验结果中，我们得到了几个有趣的观察和见解：

1) 根据测试数据集和评价指标的不同，不同方法的性能明显不同。在常用测试数据集的全参考 IQA 指标方面，MBLLEN [13]、KinD++ [61] 和 DSLR [31] 普遍优于其他比较方法。对于手机拍摄的真实世界的低照度图像，基于监督学习的 Retinex-Net [14] 和 KinD++ [61] 在非参考 IQA 指标上获得了更好的分数。对于现实世界中由手机拍摄的低光照视频，TBEFN [30] 更好地保留了时间上的一致性。在计算效率方面，LightenNet [15] 和 Zero-DCE [38] 表现出色。从黑暗中的人脸检测方面来看，TBEFN [30]、Retinex-Net [14] 和 Zero-DCE [38] 排在前三位。没有哪种方法总是胜出。总的来说，Retinex-Net [14], [30], Zero-DCE [38], 和 DSLR [31] 在大多数情况下是更好的选择。对于 raw 数据，EEMEFN [26] 比 SID [85] 获得相对更好的性能。然而，从视觉结果来看，EEMEFN [26] 和比 SID [85] 与相应的 ground truth 相比，不能很好地恢复颜色。

2)LLIV-Phone 数据集使大多数方法黯然失色。这样意味着现有方法的泛化能力需要进一步提高。值得注意的是，仅仅使用平均亮度方差来评估不同方法在低光视频增强方面的表现是不够的。更加有效和全面的评估指标将引导低照度视频增强的发展走向正道。

3) 关于学习策略，监督学习在大多数情况下取得了更好的性能，但需要高计算资源和配对的训练数据。相比之下，零次学习在实际应用中更有吸引力，因为它不需要配对或非配对的训练数据。因此，基于零次学习的方法享有更好的泛化能力。然而，它们的性能在数据上的表现不如其他方法。

4) 视觉结果和 IQA 分数之间存在差距。换句话说，一个好的视觉观感并不总是产生一个好的 IQA 分数。人类感知和 IQA 分数之间的关系值得更多的研究。追求更好的视觉感知或分数取决于具体的应用。例如，为了向观察者展示结果，应该更加关注视觉感知。相比之下，当 LLIE 方法被应用于黑暗中的人脸检测时，准确性更为重要。因此，在比较不同的方法时，应该进行更全面多角度的比较。

5) 基于深度学习的 LLIE 方法对黑暗中的人脸检测是有益的。这样的结果进一步支持了增强低照度图像和视频的意义。然而，与正常光线图像中人脸检测的高准确率相比，尽管使用了 LLIE 方法，黑暗中的人脸检测准确率也极低。

6) 与基于 RGB 格式的 LLIE 方法相比，基于 raw 格式的 LLIE 方法通常能更好地恢复细节，获得更生动的色彩，并

(a) input (b) LightenNet [15] (c) Retinex-Net [14] (d) MBLLEN [13]

(e) KinD++ [61] (f) TBEFN [30] (g) DSLR [31] (h) EnlightenGAN [36]

(i) DRBN [43] (j) ExCNet [37] (k) Zero-DCE [38] (l) RRDNet [39]

图 11. 不同方法在从 DARK FACE 数据集取样的低照度图像上的视觉结果。通过放大，可以更好地看到人脸的边界框。

更有效地减少噪音和伪影。这是因为 raw 数据包含更多的信息，如更宽的色域和更高的动态范围。然而，基于 raw 格式的 LLIE 方法仅限于特定的传感器和格式，如索尼相机的拜耳模式和富士相机的 APS-C X-Trans 模式。相比之下，基于 RGB 格式的 LLIE 方法更加方便和通用，因为 RGB 图像通常是由移动设备产生的最终图像形式。然而，基于 RGB 格式的 LLIE 方法不能很好地应对表现出低光和过度噪音的情况。

5 开放性问题

在这一节中，我们总结了低照度图像和视频增强中的开放性问题如下。

泛化能力. 尽管现有的方法可以产生一些视觉上讨喜的结果，但它们的泛化能力有限。例如，在 MIT-Adobe FiveK [79] 数据集上训练的方法不能有效地增强 LOL [14] 数据集的低照度图像。尽管合成数据被用来增加训练数据的多样性，但在真实和合成数据的组合上训练的模型不能很好地解决这个问题。提高 LLIE 方法的泛化能力是一个尚未解决的问题。

消除未知噪声. 观察现有方法对不同类型手机摄像头拍摄的低照度图像的处理结果，我们可以发现，这些方法不能很好地去除噪声，甚至会放大噪声，尤其是在噪声类型未知的情况下。尽管有些方法在其训练数据中加入了高斯和/或泊松噪声，但这些噪声类型与真实的噪声不同，因此这些方法在真实场景中的表现并不令人满意。去除未知噪声的问题仍未解决。

移除未知伪影. 人们可能会希望加强从互联网上传和下载的低照度图像。该图像可能已经经历了一系列的退化，如 JPEG 压缩或编辑。因此，该图像可能包含未知的伪影。抑制未知的伪影仍然是对现有低照度图像和视频增强方法的挑战。

纠正不均匀光照. 在真实场景中拍摄的图像通常表现出不均匀的光照。例如，在夜间拍摄的图像既有黑暗区域，也有正常光线，或过度曝光的区域如光源区域。现有的方法倾向于同时提亮黑暗区域和光源区域，这影响了增强结果的视觉质量。实际上我们更希望增强暗部区域，但抑制过度曝光区域。然而，现有的 LLIE 方法并没有很好地解决这个开放性问题。

区分语义区. 现有的方法倾向于增强低照度图像而不考虑其不同区域的语义信息。例如，在低照度图像中，一个人的黑发被增强为偏白色，因为黑发被当作低照度区域对待。一个理想的增强方法应该是只增强由外部环境引起的低光区域。如何区分语义区域是一个开放的问题。

使用相邻帧. 尽管已经提出了一些增强低照度视频的方法，但它们通常是逐帧地处理视频。如何充分利用邻近的帧来提高增强性能并加快处理速度是一个尚未解决的问题。例如，邻近帧的良好照明区域可以被用来增强当前帧。另一个例子是，处理相邻帧的估算参数可以重新用于增强当前帧以减少参数估算的时间。

6 未来研究方向

弱光增强是一个具有挑战性的研究课题。从第 4章节中的实验和第 5章节中未解决的问题可以看出，现有的方法仍有改进的余地。我们建议未来潜在的研究方向如下。

高效的学习策略. 如前所述，目前的 LLIE 模型主要采用监督学习，需要大量的配对训练数据，并可能在特定的数据集上过度拟合。尽管一些研究人员试图将无监督学习引入 LLIE，但 LLIE 和这些学习策略之间的内在关系并不清楚，它们在 LLIE 中的有效性需要进一步改进。零次学习对真实场景显示出强大的性能，同时不需要成对的训练数据。这一独特的优势表明零点学习是一个潜在的研究方向，特别是在零参考损失、深度预设和优化策略的制定上。

专项网络结构. 网络结构可以显著影响增强性能。正如之前所分析的，大多数 LLIE 深度模型采用 U-Net 或类似 U-Net 的结构。尽管它们在某些情况下取得了可喜的性能，但仍然缺乏对这种编码器-解码器网络结构是否最适合于 LLIE 任务的调查。由于一些网络结构的参数空间较大，因此需要占用较多的内存和较长的推理时间。这样的网络结构在实际应用中是不可接受的。因此，考虑到非均匀光照、小像素值、噪声抑制和颜色恒定等低照度图像的特点，研究一种更有效的网络结构用于 LLIE 是有必要的。人们还可以通过考虑低照度图像的局部相似性或考虑更有效的操作，如深度可分离卷积层（depthwise separable convolution layer）[95] 和自校准卷积（self-calibrate）[96] 来设计更有效的网络结构。可以考虑采用神经结构搜索（NAS）技术 [97], [98] 来获得更有效和高效的 LLIE 网络结构。将 Transformer 结构 [99], [100] 适配于 LLIE 可能是一个潜在的、有趣的研究方向。

损失函数. 损失函数约束了输入图像和 ground truth 之间的关系，并驱动了深度网络的优化。在 LLIE 中，常用的损失函数是从相关的视觉任务中借用的。因此，设计更适合于 LLIE 的损失函数是需要的。最近的研究表明，有可能使用深度神经网络来近似人类对图像质量的视觉感知 [101], [102]。这些想法和基本理论可以用来指导低照度增强网络的损失函数的设计。

现实训练数据. 虽然已存在几个 LLIE 的训练数据集，但其真实性、规模和多样性都落后于真实的低光场景。因此，正如章节4中所示，当前的 LLIE 深度模型在遇到真实世界场景中拍摄的低光图像时，无法达到令人满意的性能。应需要更多的努力来研究收集大规模和多样化的真实世界配对的 LLIE 训练数据集，或者产生更真实的合成数据。

标准测试数据. 目前，还没有公认的 LLIE 评估基准。研究人员倾向于选择他们的测试数据，这可能会对他们提出的方法产生偏袒。尽管有些研究者分出了一些配对数据作为测试数据，但文献中对训练和测试分区的划分大多是将就的。因此，在不同的方法之间进行公平的比较往往是不容易的，甚至是不可能的。此外，一些测试数据要么容易被增强，要么原本就不是为低光增强而收集的。我们希望有一个标准的低照度图像和视频测试数据集，其中包括大量的测试样本和相应的 ground truths，涵盖不同的场景和具有挑战性的光照条件。

特定任务的评价指标. LLIE 中普遍采用的评价指标可以在一定程度上反映图像质量。然而，如何衡量一个 LLIE 方法所增强的结果有多好，仍然是对当前 IQA 指标的挑战，特别是对于非参考测量。目前的 IQA 指标要么关注人类的视觉感知，如主观质量，要么强调机器的感知，如对高级视觉任务的影响。因此，希望在这个研究方向上会有更多的工作，努力为 LLIE 设计更准确和特定任务的评价指标。

鲁棒的泛化能力. 通过观察真实世界测试数据的实验结果发现，大多数方法由于其有限的泛化能力而表现不如人意。泛化能力差是由多个因素造成的，如合成训练数据、小规模的训练数据、无效的网络结构或不现实的假设。探索提高泛化能力的方法是很重要的。

向低照度视频增强延伸. 与其他底层视觉任务中，如视频去模糊 [103]、[104] 和视频超分辨率 [105] 视频增强的快速发展不同，低照度视频增强受到的关注较少。若将现有的 LLIE 方法直接应用于视频，往往会导致不满意的结果和闪烁的伪影。未来将需要更多的努力来有效地去除视觉闪烁，利用相邻帧之间的时间信息，并加快增强速度。

整合语义信息. 语义信息对低照度增强至关重要。它指导网络在增强过程中区分不同的区域。没有获得语义先验的网络很容易偏离一个区域的原始颜色，例如，在增强后将黑发变成灰色。因此，将语义先验因素整合到 LLIE 模型中是一个很有前途的研究方向。类似的工作已经应用在了图像超分辨率 [106], [107] 和人脸修复 [108] 方面。

致谢

本研究得到了 RIE2020 产业联盟基金产业合作项目（IAF-ICP）资助计划的支持，以及产业伙伴的现金和实物捐助。它还得到了 NTU SUG 和 NAP 的部分支持。郭春乐由 CAAI-华为 MindSpore 开放基金赞助。

参考文献

[1] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *TCE*, vol. 53, no. 4, pp. 1752–1758, 2007.

[2] M. Abdullah-AI-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *TCE*, vol. 53, no. 2, pp. 593–600, 2007.

[3] S. Wang, J. Zheng, H. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *TIP*, vol. 22, no. 9, pp. 3538–3548, 2013.

[4] X. Fu, Y. Liao, D. Zeng, Y. Huang, X. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *TIP*, vol. 24, no. 12, pp. 4965–4977, 2015.

[5] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *TIP*, vol. 26, no. 2, pp. 982–993, 2016.

[6] S. Park, S. Yu, B. Moon, S. Ko, and J. Paik, "Low-light image enhancement using variational optimization-based retinex model," *TCE*, vol. 63, no. 2, pp. 178–184, 2017.

[7] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *TIP*, vol. 27, no. 6, pp. 2828–2841, 2018.

[8] Z. Gu, F. Li, F. Fang, and G. Zhang, "A novel retinex-based fractional-order variational model for images with severely low light," *TIP*, vol. 29, pp. 3239–3253, 2019.

[9] X. Ren, W. Yang, W.-H. Cheng, and J. Liu, "LR3M: Robust low-light enhancement via low-rank regularized retinex model," *TIP*, vol. 29, pp. 5862–5876, 2020.

[10] S. Hao, X. Han, Y. Guo, X. Xu, and M. Wang, "Low-light image enhancement with semi-decoupled decomposition," *TMM*, vol. 22, no. 12, pp. 3025–3038, 2020.

[11] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *PR*, vol. 61, pp. 650–662, 2017.

[12] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *CVPR*, 2018, pp. 3291–3300.

[13] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using cnns," in *BMVC*, 2018.

[14] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *BMVC*, 2018.

[15] C. Li, J. Guo, F. Porikli, and Y. Pang, "LightenNet: A convolutional neural network for weakly illuminated image enhancement," *PRL*, vol. 104, pp. 15–22, 2018.

[16] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *TIP*, vol. 27, no. 4, pp. 2049–2062, 2018.

[17] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement usign deep illumination estimation," in *CVPR*, 2019, pp. 6849–6857.

[18] C. Chen, Q. Chen, M. N. Do, and V. Koltun, "Seeing motion in the dark," in *ICCV*, 2019, pp. 3185–3194.

[19] H. Jiang and Y. Zheng, "Learning to see moving object in the dark," in *ICCV*, 2019, pp. 7324–7333.

[20] Y. Wang, Y. Cao, Z. Zha, J. Zhang, Z. Xiong, W. Zhang, and F. Wu, "Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement," in *ACMMM*, 2019, pp. 2015–2023.

[21] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *ACMMM*, 2019, pp. 1632–1640.

[22] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang, "Low-light image enhancement via a deep hybrid network," *TIP*, vol. 28, no. 9, pp. 4364–4375, 2019.

[23] K. Xu, X. Yang, B. Yin, and R. W. H. Lau, "Learning to restore low-light images via decomposition-and-enhancement," in *CVPR*, 2020, pp. 2281–2290.

[24] M. Fan, W. Wang, W. Yang, and J. Liu, "Integrating semantic segmentation and retinex model for low light image enhancement," in *ACMMM*, 2020, pp. 2317–2325.

[25] F. Lv, B. Liu, and F. Lu, "Fast enhancement for non-uniform illumination images using light-weight cnns," in *ACMMM*, 2020, pp. 1450–1458.

[26] M. Zhu, P. Pan, W. Chen, and Y. Yang, "EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network," in *AAAI*, 2020, pp. 13 106–13 113.

[27] D. Triantafyllidou, S. Moran, S. McDonagh, S. Parisot, and G. Slabaugh, "Low light video enhancement using synthetic data produced with an intermediate domain mapping," in *ECCV*, 2020, pp. 103–119.

[28] J. Li, J. Li, F. Fang, F. Li, and G. Zhang, "Luminance-aware pyramid network for low-light image enhancement," *TMM*, 2020.

[29] L. Wang, Z. Liu, W. Siu, and D. P. K. Lun, "Lightening network for low-light image enhancement," *TIP*, vol. 29, pp. 7984–7996, 2020.

[30] K. Lu and L. Zhang, "TBEFN: A two-branch exposure-fusion network for low-light image enhancement," *TMM*, 2020.

[31] S. Lim and W. Kim, "DSLR: Deep stacked laplacian restorer for low-light image enhancement," *TMM*, 2020.

[32] F. Zhang, Y. Li, S. You, and Y. Fu, "Learning temporal consistency for low light video enhancement from single images," in *CVPR*, 2021.

[33] J. Li, X. Feng, and Z. Hua, "Low-light image enhancement via progressive-recursive network," *TCSVT*, 2021.

[34] W. Yang, W. Wang, H. Huang, S. Wang, and J. Liu, "Sparse gradient regularized deep retinex network for robust low-light image enhancement," *TIP*, vol. 30, pp. 2072–2086, 2021.

[35] R. Yu, W. Liu, Y. Zhang, Z. Qu, D. Zhao, and B. Zhang, "DeepExposure: Learning to expose photos with asynchronously reinforced adversarial learning," in *NeurIPS*, 2018, pp. 2149–2159.

[36] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "EnlightenGAN: Deep light enhancement without paired supervision," *TIP*, vol. 30, pp. 2340–2349, 2021.

[37] L. Zhang, L. Zhang, X. Liu, Y. Shen, S. Zhang, and S. Zhao, "Zero-shot restoration of back-lit images using deep internal learning," in *ACMMM*, 2019, pp. 1623–1631.

[38] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *CVPR*, 2020, pp. 1780–1789.

[39] A. Zhu, L. Zhang, Y. Shen, Y. Ma, S. Zhao, and Y. Zhou, "Zero-shot restoration of underexposed images via robust retinex decomposition," in *ICME*, 2020, pp. 1–6.

[40] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *TPAMI*, 2021.

[41] Z. Zhao, B. Xiong, L. Wang, Q. Ou, L. Yu, and F. Kuang, "Retinexdip: A unified deep framework for low-light image enhancement," *TCSVT*, 2021.

[42] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *CVPR*, 2021.

[43] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement," in *CVPR*, 2020, pp. 3063–3072.

[44] W. Yang, S. Wang, Y. F. nd Yue Wang, and J. Liu, "Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality," *TIP*, vol. 30, pp. 3461–3473, 2021.

[45] Y. Hu, H. He, C. Xu, B. Wang, and S. Lin, "Exposure: A white-box photo post-processing framework," *ACM Graph.*, vol. 37, no. 2, pp. 1–17, 2018.

[46] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Graph.*, vol. 36, no. 4, pp. 1–12, 2017.

[47] Y. Chen, Y. Wang, M. Kao, and Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement form photographs wigh gans," in *CVPR*, 2018, pp. 6306–6314.

[48] Y. Deng, C. C. Loy, and X. Tang, "Aesthetic-driven image enhancement by adversarial learning," in *ACMMM*, 2018, pp. 870–878.

[49] Z. Yan, H. Zhang, B. Wang, S. Paris, and Y. Yu, "Automatic photo adjustment using deep neural networks," *ACM Graph.*, vol. 35, no. 2, pp. 1–15, 2016.

[50] Q. Chen, J. Xu, and V. Koltun, "Fast image processing with fully-convolutional networks," in *CVPR*, 2017, pp. 2497–2506.

[51] Z. Ni, W. Yang, S. Wang, L. Ma, and S. Kwong, "Towards unsupervised deep image enhancement with generative adversarial network," *TIP*, vol. 29, pp. 9140–9151, 2020.

[52] H. Zeng, J. Cai, L. Li, Z. Cao, and L. Zhang, "Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time," *TPAMI*, 2020.

[53] C. Li, C. Guo, Q. Ai, S. Zhou, and C. C. Loy, "Flexible piecewise curves estimation for photo enhancement," *arXiv preprint arXiv:2010.13412*, 2020.

[54] W. Wang, X. Wu, X. Yuan, and Z. Gao, "An experiment-based review of low-light image enhancement methods," *IEEE Access*, vol. 8, pp. 87 884–87 917, 2020.

[55] J. Liu, D. Xu, W. Yang, M. Fan, and H. Huang, "Benchmarking low-light image enhancement and beyond," *IJCV*, 2021.

[56] V. Jain and S. Seung, "Natural image denoising with convolutional networks," in *NeurIPS*, 2008, pp. 1–8.

[57] K. Xu, X. Yang, B. Yin, and R. W. H. Lau, "Learning to restore low-light images via decomposition-and-enhancement," in *CVPR*, 2020, pp. 2281–2290.

[58] E. H. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *National Academy of Sciences*, vol. 83, no. 10, pp. 3078–3080, 1986.

[59] D. J. Jobson, Z. ur Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *TIP*, vol. 6, no. 3, pp. 451–462, 1997.

[60] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *CVPR*, 2019, pp. 6849–6857.

[61] X. Guo, Y. Zhang, J. Ma, W. Liu, and J. Zhang, "Beyond brightening low-light images," *IJCV*, 2020.

[62] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[63] T. Xue, B. Chen, J. Wu, D. Wei, and W. T. Freeman, "Video enhancement with task-oriented flow," *IJCV*, vol. 127, no. 8, pp. 1106–1125, 2019.

[64] K. He, J. Sun, and X. Tang, "Guided image filtering," *TPAMI*, vol. 35, no. 6, pp. 1397–1409, 2013.

[65] P. Whittle, "The psychophysics of contrast brightness," *A. L. Gilchrist (Ed.), Brightness, lightness, and transparenc (1994),*, pp. 35–110, 1993.

[66] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *CVPR*, 2018.

[67] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *TIP*, vol. 3, no. 1, pp. 47–56, 2017.

[68] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *CVPR*, 2017, pp. 4681–4690.

[69] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *CVPR*, 2009, pp. 248–255.

[70] K. Simoayan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[71] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *CVPR*, 2017.

[72] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *ECCVW*, 2018.

[73] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *TPAMI*, vol. 38, no. 2, pp. 295–307, 2015.

[74] Q. Xu, C. Zhang, and L. Zhang, "Denoising convolutional neural network," in *ICIA*, 2015, pp. 1184–1187.

[75] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *CVPR*, 2017, pp. 3855–3863.

[76] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Deep joint rain detection and removal from a single image," in *CVPR*, 2017, pp. 1357–1366.

[77] X. Fu, Q. Qi, Z. Zha, X. Ding, F. Wu, and J. Paisley, "Successive graph convolutional network for image deraining," *IJCV*, vol. 129, no. 5, pp. 1691–1711, 2021.

[78] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *CVPR*, 2015, pp. 769–777.

[79] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *CVPR*, 2011, pp. 97–104.

[80] Y. Yuan, W. Yang, W. Ren, J. Liu, W. JScheirer, and W. Zhangyang, "UG+ Track 2: A collective benchmark effort for evaluating and advancing image understanding in poor visibility environments," *arXiv arXiv:1904.04474*, 2019.

[81] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *CVIU*, vol. 178, pp. 30–42, 2019.

[82] L. Chulwoo, L. Chul, L. Young-Yoon, and K. Chang-su, "Power-constrained contrast enhancement for emissive displays based on histogram equalization," *TIP*, vol. 21, no. 1, pp. 80–93, 2012.

[83] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *TIP*, vol. 22, no. 12, pp. 5372–5384, 2013.

[84] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving video database with scalable annotation tooling," *arXiv preprint arXiv:1805.04687*, 2018.

[85] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *CVPR*, 2018, pp. 3291–3300.

[86] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *TIP*, vol. 13, no. 4, pp. 600–612, 2004.

[87] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018, pp. 586–595.

[88] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *SPL*, vol. 20, no. 3, pp. 209–212, 2013.

[89] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *CVPR*, 2018, pp. 6228–6237.

[90] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a non-reference quality metric for single-image super-resolution," *CVIU*, vol. 158, pp. 1–16, 2017.

[91] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang, "Perceptual quality assessment of smartphone photography," in *ICCV*, 2020, pp. 3677–3686.

[92] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: Efficient convolution operators for tracking," in *CVPR*, 2017, pp. 6638–6646.

[93] S. Yang, P. Luo, C. C. Loy, and X. Tang, "Wider Face: A face detection benchmark," in *CVPR*, 2016, pp. 5525–5533.

[94] J. Li, Y. Wang, C. Wang, Y. Tai, J. Qian, J. Yang, C. Wang, J. Li, and F. Huang, "DSFD: Dual shot face detector," in *CVPR*, 2019, pp. 5060–5069.

[95] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision application," *arXiv preprint arXiv:1704.04861*, 2017.

[96] J. Liu, Q. Hou, M. M. Cheng, C. Wang, and J. Feng, "Improving convolutional networks with self-calibrated convolutions," in *CVPR*, 2020, pp. 10 096–10 105.

[97] C. Liu, B. Zoph, M. Neumann, J. Shlens, W. Hua, L. Li, L. F. Fei, A. Yuille, J. Huang, and K. Murphy, "Progressive neural architecture search," in *ECCV*, 2018, pp. 19–34.

[98] C. Liu, L. C. Chen, F. Schroff, H. Adam, W. Hua, A. Yuille, and L. F. Fei, "Auto-Deeplab: Hierarchical neural architecture search for semantic image segmentation," in *CVPR*, 2019, pp. 82–92.

[99] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. D. M. Minderer, G. Heigold, S. Gelly, J. Uszekoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[100] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, "Pre-trained image processing transformer," *arXiv preprint arXiv:2012.00364*, 2020.

[101] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang, "Perceptual quality assessment of smartphone phtography," in *CVPR*, 2020, pp. 3677–3686.

[102] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *TIP*, vol. 27, no. 8, pp. 3998–4011, 2018.

[103] T. H. Kim, K. M. Lee, B. Scholkopf, and M. Hirsch, "Online video deblurring via dynamic temporal blending network," in *ICCV*, 2017, pp. 4038–4047.

[104] T. Ehret, A. Davy, J.-M. Morel, G. Facciolo, and P. Arias, "Model-blind video denoising via frame-to-frame training," in *CVPR*, 2019, pp. 11 369–11 378.

[105] K. C. K. Chan, X. Wang, K. Yu, C. Dong, and C. C. Loy, "BasicVSR: The search for essential components in video super-resolution and beyond," in *CVPR*, 2021.

[106] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *CVPR*, 2018, pp. 606–615.

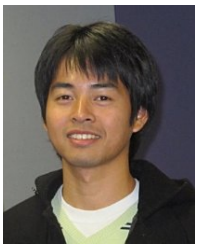
[107] K. C. K. Chan, X. Wang, X. Xu, J. Gu, and C. C. Loy, "GLEAN: Generative latent bank for large-factor image super-resolution," in *CVPR*, 2021.

[108] X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo, and L. Zhang, "Blind face restoration via deep multi-scale component dictionaries," in *ECCV*, 2020, pp. 399–415.

Chongyi Li is a Research Assistant Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He received the Ph.D. degree from Tianjin University, China in 2018. From 2016 to 2017, he was a joint-training Ph.D. Student with Australian National University, Australia. Prior to joining NTU, he was a postdoctoral fellow with City University of Hong Kong and Nanyang Technological University from 2018 to 2021. His current research focuses on image processing, computer vision, and deep learning, particularly in the domains of image restoration and enhancement. He serves as an associate editor of the Journal of Signal, Image and Video Processing and a lead guest editor of the IEEE Journal of Oceanic Engineering.

Chunle Guo received his PhD degree from Tianjin University in China under the supervision of Prof. Jichang Guo. He conducted the Ph.D. research as a Visiting Student with the School of Electronic Engineering and Computer Science, Queen Mary University of London (QMUL), UK. He continued his research as a Research Associate with the Department of Computer Science, City University of Hong Kong (CityU), from 2018 to 2019. Now he is a postdoc research fellow working with Prof. Ming-Ming Cheng at Nankai University. His research interests lie in image processing, computer vision, and deep learning.

Linhao Han is currently a master student at the College of Computer Science, Nankai University, under the supervision of Prof. Ming-Ming Cheng. His research interests include deep learning and computer vision.



Jun Jiang received the PhD degree in Color Science from Rochester Institute of Technology in 2013. He is a Senior Researcher in SenseBrain focusing on algorithm development to improve image quality on smartphone cameras. His research interest includes computational photography, low-level computer vision, and deep learning.

Ming-Ming Cheng (Senior Member, IEEE) received the Ph.D. degree from Tsinghua University in 2012. Then he did two years research fellowship with Prof. Philip Torr at Oxford. He is currently a Professor at Nankai University and leading the Media Computing Laboratory. His research interests include computer graphics, computer vision, and image processing. He received research awards, including the ACM China Rising Star Award, the IBM Global SUR Award, and the CCF-Intel Young Faculty Researcher Program. He is on the Editorial Board Member of IEEE Transactions on Image Processing (TIP).

Jinwei Gu (Senior Member, IEEE) is the R&D Executive Director of SenseTime USA. His current research focuses on low-level computer vision, computational photography, smart visual sensing and perception, and robotics. He obtained his Ph.D. degree in 2010 from Columbia University, and his B.S and M.S. from Tsinghua University, in 2002 and 2005 respectively. Before joining SenseTime, he was a senior research scientist in NVIDIA Research from 2015 to 2018. Prior to that, he was an assistant professor in Rochester Institute of Technology from 2010 to 2013, and a senior researcher in the media lab of Futurewei Technologies from 2013 to 2015. He is an associate editor for IEEE Transactions on Computational Imaging and an IEEE senior member since 2018.

Chen Change Loy (Senior Member, IEEE) is an Associate Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He is also an Adjunct Associate Professor at The Chinese University of Hong Kong. He received his Ph.D. (2010) in Computer Science from the Queen Mary University of London. Prior to joining NTU, he served as a Research Assistant Professor at the MMLab of The Chinese University of Hong Kong, from 2013 to 2018. He was a postdoctoral researcher at Queen Mary University of London and Vision Semantics Limited, from 2010 to 2013. He serves as an Associate Editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence and International Journal of Computer Vision. He also serves/served as an Area Chair of major conferences such as ICCV, CVPR, ECCV and AAAI. His research interests include image/video restoration and enhancement, generative tasks, and representation learning.